

High-Dimension, Low Sample Size Asymptotics of Canonical Correlation Analysis

Sungwon Lee, Statistics Department, University of Pittsburgh

Abstract

An asymptotic behavior of canonical correlation analysis is studied when dimension d grows and the sample size n is fixed. In particular, we are interested in the conditions for which CCA works or fails in the HDLSS situation. This technical report investigates those conditions in a rather simplified setting where there exists one pair of directions in two sets of random variables with non-zero correlation between two sets of scores on them. Proofs and an extensive simulation study supports the findings.

1 Introduction

Canonical correlation analysis (CCA) introduced in [2] is a standard statistical tool to explore the relationship between two sets of random variables. Consider d_X - and d_Y -dimensional random vectors $X^{(d_X)}$ and $Y^{(d_Y)}$,

$$\left(X^{(d_X)}\right)^T = \begin{bmatrix} X_1, & X_2, & \dots, & X_{d_X} \end{bmatrix}, \quad \left(Y^{(d_Y)}\right)^T = \begin{bmatrix} Y_1, & Y_2, & \dots, & Y_{d_Y} \end{bmatrix}.$$

CCA first seeks a pair of d_X - and d_Y -dimensional weights vectors $\psi_{X1}^{(d_X)}$ and $\psi_{Y1}^{(d_Y)}$ such that two random variables, one being the linear combination of X_1, X_2, \dots, X_{d_X} weighted by the elements of $\psi_{X1}^{(d_X)}$ and the other being that of Y_1, Y_2, \dots, Y_{d_Y} weighted by the elements of $\psi_{Y1}^{(d_Y)}$, have a maximal correlation,

$$(\psi_{X1}^{(d_X)}, \psi_{Y1}^{(d_Y)}) = \underset{\text{Var}(\langle \psi_{X1}^{(d_X)}, X^{(d_X)} \rangle) = \text{Var}(\langle \psi_{Y1}^{(d_Y)}, Y^{(d_Y)} \rangle) = 1}{\text{argmax}} \text{Cov}(\langle \psi_{X1}^{(d_X)}, X^{(d_X)} \rangle, \langle \psi_{Y1}^{(d_Y)}, Y^{(d_Y)} \rangle). \quad (1)$$

Requiring the norms of the weight vectors $\psi_{X1}^{(d_X)}$ and $\psi_{Y1}^{(d_Y)}$ to be one, the equation (1) can be written as an equivalent form of,

$$(\psi_{X1}^{(d_X)}, \psi_{Y1}^{(d_Y)}) = \underset{\|\psi_{X1}^{(d_X)}\|_2 = \|\psi_{Y1}^{(d_Y)}\|_2 = 1}{\operatorname{argmax}} \frac{\operatorname{Cov}(\langle \psi_{X1}^{(d_X)}, X^{(d_X)} \rangle, \langle \psi_{Y1}^{(d_Y)}, Y^{(d_Y)} \rangle)}{\sqrt{\operatorname{Var}(\langle \psi_{X1}^{(d_X)}, X^{(d_X)} \rangle)} \sqrt{\operatorname{Var}(\langle \psi_{Y1}^{(d_Y)}, Y^{(d_Y)} \rangle)}}. \quad (2)$$

For convenience, denote the objective function in the right hand side of (2) by $\rho_P(\psi^{(d_X)}, \psi^{(d_Y)})$,

$$\begin{aligned} \rho : R^{d_X} \times R^{d_Y} &\mapsto R \\ \rho_P(\psi^{(d_X)}, \psi^{(d_Y)}) &= \frac{\operatorname{Cov}(\langle \psi^{(d_X)}, X^{(d_X)} \rangle, \langle \psi^{(d_Y)}, Y^{(d_Y)} \rangle)}{\sqrt{\operatorname{Var}(\langle \psi^{(d_X)}, X^{(d_X)} \rangle)} \sqrt{\operatorname{Var}(\langle \psi^{(d_Y)}, Y^{(d_Y)} \rangle)}}. \end{aligned}$$

Subsequent weights vectors $\psi_{Xi}^{(d_X)}$ and $\psi_{Yi}^{(d_Y)}$, for $i = 1, 2, \dots, \min(d_X, d_Y)$, are found by maximizing the objective function $\rho_P(\psi^{(d_X)}, \psi^{(d_Y)})$,

$$(\psi_{Xi}^{(d_X)}, \psi_{Yi}^{(d_Y)}) = \underset{\|\psi_{Xi}^{(d_X)}\|_2 = \|\psi_{Yi}^{(d_Y)}\|_2 = 1}{\operatorname{argmax}} \rho_P(\psi_{Xi}^{(d_X)}, \psi_{Yi}^{(d_Y)}), \quad i = 1, 2, \dots, \min(d_X, d_Y),$$

under the constraint that,

$$\begin{aligned} \operatorname{Cov}(\langle \psi_{Xi}^{(d_X)}, X^{(d_X)} \rangle, \langle \psi_{Xj}^{(d_X)}, X^{(d_X)} \rangle) &= \operatorname{Cov}(\langle \psi_{Yi}^{(d_Y)}, Y^{(d_Y)} \rangle, \langle \psi_{Yj}^{(d_Y)}, Y^{(d_Y)} \rangle) \\ &= \operatorname{Cov}(\langle \psi_{Xi}^{(d_X)}, X^{(d_X)} \rangle, \langle \psi_{Yj}^{(d_Y)}, Y^{(d_Y)} \rangle) \\ &= \operatorname{Cov}(\langle \psi_{Yi}^{(d_Y)}, Y^{(d_Y)} \rangle, \langle \psi_{Xj}^{(d_X)}, X^{(d_X)} \rangle) \\ &= 0, \quad i = 1, 2, \dots, \min(d_X, d_Y), \quad j = 1, 2, \dots, i-1 \text{ for each } i. \end{aligned}$$

The i th pair of weight vectors $\psi_{Xi}^{(d_X)}$ and $\psi_{Yi}^{(d_Y)}$ are usually called the i th pair of canonical weight vectors (or canonical loadings). The correlation ρ evaluated at the i th pair $\psi_{Xi}^{(d_X)}$ and $\psi_{Yi}^{(d_Y)}$, denoted by $\rho_i^{(d_X, d_Y)}$, is called the i th canonical correlation coefficient, that is, $\rho_i^{(d_X, d_Y)} = \rho_P(\psi_{Xi}^{(d_X)}, \psi_{Yi}^{(d_Y)})$.

In practice, we collect two sets of observations of d_X - and d_Y -dimensional random vectors $X^{(d_X)}$ and $Y^{(d_Y)}$ on a common set of samples in a $d_X \times n$ matrix $\mathbf{X}^{(d_X)}$ and a $d_Y \times n$ matrix $\mathbf{Y}^{(d_Y)}$, respectively. We row-center $\mathbf{X}^{(d_X)}$ and $\mathbf{Y}^{(d_Y)}$ and let $\hat{\Sigma}_X^{(d_X)}$, $\hat{\Sigma}_Y^{(d_Y)}$ and $\hat{\Sigma}_{XY}^{(d_X, d_Y)}$ be a covariance

matrix of $X^{(d_X)}$, a covariance matrix of $Y^{(d_Y)}$ and a cross-covariance matrix of $X^{(d_X)}$ and $Y^{(d_Y)}$,

$$\hat{\Sigma}_X^{(d_X)} = \frac{1}{n} \mathbf{X}^{(d_X)} \left(\mathbf{X}^{(d_X)} \right)^T, \quad \hat{\Sigma}_Y^{(d_Y)} = \frac{1}{n} \mathbf{Y}^{(d_Y)} \left(\mathbf{Y}^{(d_Y)} \right)^T, \quad \hat{\Sigma}_{XY}^{(d_X, d_Y)} = \frac{1}{n} \mathbf{X}^{(d_X)} \left(\mathbf{Y}^{(d_Y)} \right)^T.$$

For the case where the sample size n is greater than d_X and d_Y , the estimation of sample canonical weight vectors $(\hat{\psi}_{Xi}^{(d_X)}, \hat{\psi}_{Yi}^{(d_Y)})$ and sample canonical correlation coefficients $\hat{\rho}_i^{(d_X, d_Y)}$ are done through singular value decomposition of the matrix $\hat{\mathbf{R}}^{(d_X, d_Y)}$,

$$\begin{aligned} \hat{\mathbf{R}}^{(d_X, d_Y)} &= \left(\hat{\Sigma}_X^{(d_X)} \right)^{-\frac{1}{2}} \hat{\Sigma}_{XY}^{(d_X, d_Y)} \left(\hat{\Sigma}_Y^{(d_Y)} \right)^{-\frac{1}{2}}, \\ \text{SVD}(\hat{\mathbf{R}}^{(d_X, d_Y)}) &= \sum_{i=1}^{\min(d_X, d_Y)} \hat{\lambda}_{Ri}^{(d_X, d_Y)} \hat{\eta}_{RXi}^{(d_X)} \left(\hat{\eta}_{RYi}^{(d_Y)} \right)^T, \end{aligned} \quad (3)$$

where $\hat{\lambda}_{Ri}^{(d)}$ is a sample singular value with $\hat{\lambda}_{R1}^{(d)} \geq \hat{\lambda}_{R2}^{(d)} \geq \dots \geq \hat{\lambda}_{R\min(d_X, d_Y)}^{(d)} \geq 0$, and $(\hat{\eta}_{RXi}^{(d_X)}, \hat{\eta}_{RYi}^{(d_Y)})$ is a pair of left and right sample singular vectors corresponding to $\hat{\lambda}_{Ri}^{(d)}$. Then, the i th sample canonical correlation coefficient $\hat{\rho}_i^{(d)}$ is found to be,

$$\hat{\rho}_i^{(d_X, d_Y)} = \hat{\lambda}_{Ri}^{(d_X, d_Y)}.$$

The i th pair of canonical weight vectors $\hat{\psi}_{Xi}^{(d_X)}$ and $\hat{\psi}_{Yi}^{(d_Y)}$ are obtained by unscaling and normalizing the i th pair of sample singular vectors $\hat{\eta}_{RXi}^{(d_X)}$ and $\hat{\eta}_{RYi}^{(d_Y)}$,

$$\hat{\psi}_{Xi}^{(d_X)} = \frac{\left(\hat{\Sigma}_X^{(d_X)} \right)^{-\frac{1}{2}} \hat{\eta}_{RXi}^{(d_X)}}{\left\| \left(\hat{\Sigma}_X^{(d_X)} \right)^{-\frac{1}{2}} \hat{\eta}_{RXi}^{(d_X)} \right\|_2}, \quad \hat{\psi}_{Yi}^{(d_Y)} = \frac{\left(\hat{\Sigma}_Y^{(d_Y)} \right)^{-\frac{1}{2}} \hat{\eta}_{RYi}^{(d_Y)}}{\left\| \left(\hat{\Sigma}_Y^{(d_Y)} \right)^{-\frac{1}{2}} \hat{\eta}_{RYi}^{(d_Y)} \right\|_2}. \quad (4)$$

The projection of the data matrix $\mathbf{X}^{(d_X)}$ onto the i th sample canonical weight vector $\hat{\psi}_{Xi}^{(d_X)}$ gives the canonical scores (or canonical variables) of $\mathbf{X}^{(d_X)}$ with respect to $\hat{\psi}_{Xi}^{(d_X)}$ and similarly for $\mathbf{X}^{(d_X)}$. Although powerful, CCA has several disadvantages. first, use of CCA is practically restricted to the case of two sets of data even if there is an attempt to generalize it to more than two sets of data [7]. Second, CCA components are estimable only if the sample size n is greater than d_X and d_Y . It is well know that, when $n < \max(d_X, d_Y)$, one can construct an infinite number of sample canonical weight vector pairs with their correlation of one. Moreover, overfitting is often a problem

even when $n > d_X$ and d_Y . Hence, CCA is often considered not reliable in high-dimensional data sets. We, however, will show that, even in the case where sample size n is less than d_X or d_Y , some sample canonical weight vectors is estimable and furthermore consistent under a certain condition.

As high-dimensional data are increasingly common these days, where a large number of variables are measured for each object, there is a strong need to investigate the behavior of estimates resulting from the application of standard statistical tools such as CCA to a high-dimensional case (that is, scalability of those tools). Literature in the HDLSS asymptotic study of CCA is very limited, while the behavior of PCA components under the similar high-dimensional condition is well-studied in [3, 4]. This might be in part because CCA is not as widely used as PCA, which is almost an indispensable tool for dimension reduction of high-dimensional data prevalent these days, and in part due to the complicated estimation steps involving an inverse operator as in (3), which makes the analysis not straightforward.

A relevant work is first addressed in [5], where the asymptotic behavior of sample singular vectors and singular values are analysed under a HDLSS setting. In [6], the similar study of CCA is elaborated on, but their proof should have considered the fact that an infinite sum of quantities converging to zero does not necessarily approach to zero. The HDLSS asymptotic behavior of CCA components in this chapter will be studied in relatively a simple population structure and serves as a groundwork for further analysis.

2 Assumptions and Definitions

Without loss of generality for the case where the dimensions of two random vectors $X^{(d_X)}$ and $Y^{(d_Y)}$ grow in a sense that $d_X/d_Y \rightarrow 1$, we set $d_X = d_Y$ and consider two random vectors $X^{(d)}$ and $Y^{(d)}$ of a same dimension with mean zero. We assume that covariance structure of $X^{(d)}$ and $Y^{(d)}$ follows a simple spiked model as in [1], where the leading eigenvalues of their covariance matrix is considerably larger than the rest. In specific, let $\Sigma_X^{(d)}$ and $\Sigma_Y^{(d)}$ be the covariance matrices of $X^{(d)}$ and $Y^{(d)}$. Then, a spiked model can be easily understood via eigendecomposition of $\Sigma_X^{(d)}$ and $\Sigma_Y^{(d)}$,

$$\Sigma_X^{(d)} = \sum_{i=1}^d \lambda_{X_i}^{(d)} \xi_{X_i}^{(d)} \left(\xi_{X_i}^{(d)} \right)^T, \quad \Sigma_Y^{(d)} = \sum_{j=1}^d \lambda_{Y_j}^{(d)} \xi_{Y_j}^{(d)} \left(\xi_{Y_j}^{(d)} \right)^T, \quad (5)$$

where $\lambda_{X_i}^{(d)}$ is an polpulation eigenvalue (or population PC variance) with $\lambda_{X1}^{(d)} \geq \lambda_{X2}^{(d)} \geq \dots \geq \lambda_{Xd}^{(d)} \geq 0$, $\xi_{X_i}^{(d)}$ is an population eigenvector (or population PC direction) with $\|\xi_{X_i}^{(d)}\|_2 = 1$ and $\langle \xi_{X_i}^{(d)}, \xi_{X_j}^{(d)} \rangle = 0$ for $i \neq j$ and similarly for $\lambda_{Y_j}^{(d)}$ and $\xi_{Y_j}^{(d)}$. Here, we set,

$$\begin{aligned} \lambda_{X1}^{(d)} &= \sigma_X^2 d^\alpha \text{ and } \lambda_{X_i}^{(d)} = \tau_X^2 \text{ for } i = 2, 3, \dots, d, \\ \lambda_{Y1}^{(d)} &= \sigma_Y^2 d^\alpha \text{ and } \lambda_{Y_j}^{(d)} = \tau_Y^2 \text{ for } j = 2, 3, \dots, d, \end{aligned} \quad (6)$$

where one sees that the leading eigenvalues $\lambda_{X1}^{(d)}$ and $\lambda_{Y1}^{(d)}$ become dominating the rest as $d \rightarrow \infty$. We now set up the population canonical components. We assume that the two random vector is related by a pair of canonical weight vectors with its canonical correlation coefficient of ρ . The population canonical weight vector $\psi_X^{(d)}$ in the $X^{(d)}$ part is a linear combination of two eigenvectors $\xi_{X1}^{(d)}$ and $\xi_{X2}^{(d)}$ without loss of generality ($\xi_{X2}^{(d)}$ can be replaced with $\xi_{X_i}^{(d)}$ for any i) and similarly for the other population canonical weight vector $\psi_Y^{(d)}$ in the $Y^{(d)}$ part,

$$\psi_X^{(d)} = \cos \theta_X \xi_{X1}^{(d)} + \sin \theta_X \xi_{X2}^{(d)}, \quad \psi_Y^{(d)} = \cos \theta_Y \xi_{Y1}^{(d)} + \sin \theta_Y \xi_{Y2}^{(d)}. \quad (7)$$

Note that the angle between $\psi_X^{(d)}$ and $\xi_{X1}^{(d)}$ is θ_X and that the angle between $\psi_Y^{(d)}$ and $\xi_{Y1}^{(d)}$ is θ_Y as $\langle \psi_X^{(d)}, \xi_{X1}^{(d)} \rangle = \cos \theta_X$ and $\langle \psi_Y^{(d)}, \xi_{Y1}^{(d)} \rangle = \cos \theta_Y$. At this point, we apply the change of basis to the spaces of $X^{(d)}$ and $Y^{(d)}$ so that the eigenvectors $\{\xi_{X_i}^{(d)}\}_{i=1}^d$ and $\{\xi_{Y_i}^{(d)}\}_{i=1}^d$ are represented by the standard basis $\{e_i^{(d)}\}_{i=1}^d$. Then, the canonical weight vectors $(\psi_X^{(d)}, \psi_Y^{(d)})$ given in (7) is rewritten as,

$$\psi_X^{(d)} = \cos \theta_X e_1^{(d)} + \sin \theta_X e_2^{(d)}, \quad \psi_Y^{(d)} = \cos \theta_Y e_1^{(d)} + \sin \theta_Y e_2^{(d)},$$

and the covariance structures given in (5) and (6) are described as,

$$\Sigma_X^{(d)} = \text{diag}(\sigma_X^2 d^\alpha, \tau_X^2, \tau_X^2, \dots, \tau_X^2), \quad \Sigma_Y^{(d)} = \text{diag}(\sigma_Y^2 d^\alpha, \tau_Y^2, \tau_Y^2, \dots, \tau_Y^2), \quad (8)$$

where $\text{diag}(\bullet)$ is a square matrix with entries of \bullet in the main diagonal and 0 off of it. With these population covariance structures and canonical components, the multivariate version of the

corollary ?? gives the cross-covariance structure of $X^{(d)}$ and $Y^{(d)}$ as follows,

$$\Sigma_{XY}^{(d)} = \begin{bmatrix} \frac{\rho\sigma_X^2\sigma_Y^2d^{2\alpha}\cos\theta_X\cos\theta_Y}{AB} & \frac{\rho\sigma_X^2d^\alpha\tau_Y^2\cos\theta_X\sin\theta_Y}{AB} & \mathbf{0}_{1 \times (d-2)} \\ \frac{\rho\tau_X^2\sigma_Y^2d^\alpha\sin\theta_X\cos\theta_Y}{AB} & \frac{\rho\tau_X^2\tau_Y^2\sin\theta_X\sin\theta_Y}{AB} & \mathbf{0}_{1 \times (d-2)} \\ \mathbf{0}_{(d-2) \times 1} & \mathbf{0}_{(d-2) \times 1} & \mathbf{0}_{(d-2) \times (d-2)} \end{bmatrix}, \quad (9)$$

where

$$A = \sqrt{\sigma_X^2d^\alpha\cos^2\theta_X + \tau_X^2\sin^2\theta_X}, \quad B = \sqrt{\sigma_Y^2d^\alpha\cos^2\theta_Y + \tau_Y^2\sin^2\theta_Y}.$$

Then the covariance and cross-covariance structure of $X^{(d)}$ and $Y^{(d)}$ is succinctly described by the covariance structure of the concatenated random vector $T^{(2d)}$,

$$T^{(2d)} = \begin{bmatrix} X^{(d)} \\ Y^{(d)} \end{bmatrix}, \quad \Sigma_T^{(2d)} = \begin{bmatrix} \Sigma_X^{(d)} & \Sigma_{XY}^{(d)} \\ \left(\Sigma_{XY}^{(d)}\right)^T & \Sigma_Y^{(d)} \end{bmatrix}. \quad (10)$$

To make the analysis a bit easy, we are going to work with a different representation of $X^{(d)}$ and $Y^{(d)}$. Let $Z^{(2d)}$ be the $2d$ -dimensional standard normal random vector. Then, $T^{(2d)}$ can be expressed as,

$$T^{(2d)} = \begin{bmatrix} X^{(d)} \\ Y^{(d)} \end{bmatrix} = \left(\Sigma_T^{(2d)}\right)^{\frac{1}{2}} Z^{(2d)}, \quad Z^{(2d)} \sim N\left(\mathbf{0}_{2d \times 1}, \mathbf{I}_{2d \times 2d}\right). \quad (11)$$

We state some definitions used in the estimation. Since the dimensionality d is much larger than the sample size n in the HDLSS setting, the estimation step (3) of canonical components is problematic as the sample covariance matrices $\hat{\Sigma}_X^{(d)}$ and $\hat{\Sigma}_Y^{(d)}$ are singular. There are two ways to handle this singularity situation. The first one is to add a minute perturbation of $\epsilon\mathbf{I}$ for a small $\epsilon > 0$ to $\hat{\Sigma}_X^{(d)}$ and $\hat{\Sigma}_Y^{(d)}$ and the second is to use a pseudoinverse such as Moore-Penrose pseudoinverse. We use the pseudoinverse obtained from the eigendecomposition of the sample covariance matrices,

$$\hat{\Sigma}_X^{(d)} = \sum_{i=1}^n \hat{\lambda}_{Xi}^{(d)} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)}\right)^T, \quad \hat{\Sigma}_Y^{(d)} = \sum_{j=1}^n \hat{\lambda}_{Yj}^{(d)} \hat{\xi}_{Yj}^{(d)} \left(\hat{\xi}_{Yj}^{(d)}\right)^T, \quad (12)$$

where $\hat{\lambda}_{Xi}^{(d)}$ is a sample eigenvalue (or sample PC variance) with $\hat{\lambda}_{X1}^{(d)} \geq \hat{\lambda}_{X2}^{(d)} \geq \dots \geq \hat{\lambda}_{Xd}^{(n)} \geq 0$, $\hat{\xi}_{Xi}^{(d)}$ is a sample eigenvector (or sample PC direction) with $\|\hat{\xi}_{Xi}^{(d)}\|_2 = 1$ and $\langle \hat{\xi}_{Xi}^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle = 0$ for $i \neq j$ and similarly for $\hat{\lambda}_{Yj}^{(d)}$ and $\hat{\xi}_{Yj}^{(d)}$. The pseudoinverse we employ is defined as,

$$\left(\hat{\Sigma}_X^{(d)}\right)^{-1} = \sum_{i=1}^n \left(\hat{\lambda}_{Xi}^{(d)}\right)^{-1} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)}\right)^T, \quad \left(\hat{\Sigma}_Y^{(d)}\right)^{-1} = \sum_{j=1}^d \left(\hat{\lambda}_{Yj}^{(d)}\right)^{-1} \hat{\xi}_{Yj}^{(d)} \left(\hat{\xi}_{Yj}^{(d)}\right)^T. \quad (13)$$

Then, the sample canonical correlation coefficient $\hat{\rho}_i^{(d)}$ is found as an i th sample singular value from the SVD of the matrix $\hat{R}^{(d)}$ defined in (4). The sample canonical weight vectors $\hat{\psi}_{Xi}^{(d)}$ and $\hat{\psi}_{Yi}^{(d)}$ corresponding to $\hat{\rho}_i^{(d)}$ are obtained from (4) using the pseudoinverses (13).

The success and failure of CCA can be described by the consistency of the sample canonical weight vectors $\hat{\psi}_X^{(d)}$ and $\hat{\psi}_Y^{(d)}$ with their population counterpart $\psi_X^{(d)}$ and $\psi_Y^{(d)}$ under the limiting operation of $d \rightarrow \infty$ and n fixed. Using the angle as a measure of consistency, we say that $\hat{\psi}_X^{(d)}$ (similarly $\hat{\psi}_Y^{(d)}$) is,

- Consistent with $\psi_X^{(d)}$ if $\text{angle}(\hat{\psi}_X^{(d)}, \psi_X^{(d)}) \rightarrow 0$ as $d \rightarrow \infty$,
- Inconsistent with $\psi_X^{(d)}$ if $\text{angle}(\hat{\psi}_X^{(d)}, \psi_X^{(d)}) \rightarrow a$, for $0 < a < \pi/2$, as $d \rightarrow \infty$,
- Strongly inconsistent with $\psi_X^{(d)}$ if $\text{angle}(\hat{\psi}_X^{(d)}, \psi_X^{(d)}) \rightarrow \pi/2$ as $d \rightarrow \infty$.

Strong inconsistency implies that the estimate $\hat{\psi}_X^{(d)}$ and $\hat{\psi}_Y^{(d)}$ become completely oblivious of its population structure and reduce to arbitrary quantities, as indicated in the fact that $\pi/2$ is indeed a largest angle possible between two vectors.

3 Main Theorem and Interpretation

3.1 Main theorem

Let $X^{(d)}$ and $Y^{(d)}$ be the d -dimensional random vectors from the multivariate Gaussian distributions with mean 0 and the simple spiked covariance matrices $\Sigma_X^{(d)}$ and $\Sigma_Y^{(d)}$ described in (5) and (6). With the population canonical correlation coefficient ρ for $0 \leq \rho \leq 1$, define the population canonical weight vectors $\psi_X^{(d)}$ and $\psi_Y^{(d)}$ as,

$$\psi_X^{(d)} = \cos \theta_X \xi_{X1}^{(d)} + \sin \theta_X \xi_{X2}^{(d)}, \quad \psi_Y^{(d)} = \cos \theta_Y \xi_{Y1}^{(d)} + \sin \theta_Y \xi_{Y2}^{(d)}$$

so that the angle between $\psi_X^{(d)}$ and $\xi_{X1}^{(d)}$ is θ_X , and the angle between $\psi_Y^{(d)}$ and $\xi_{Y1}^{(d)}$ is θ_Y . Then, the cross-covariance matrix $\Sigma_{XY}^{(d)}$ of $X^{(d)}$ and $Y^{(d)}$ is found as in 9. The two random variables $X^{(d)}$ and $Y^{(d)}$ can be written in a equivalent form,

$$\begin{bmatrix} X^{(d)} \\ Y^{(d)} \end{bmatrix} = \begin{bmatrix} \Sigma_X^{(d)} & \Sigma_{XY}^{(d)} \\ \left(\Sigma_{XY}^{(d)}\right)^T & \Sigma_Y^{(d)} \end{bmatrix} Z^{(2d)}, \quad (14)$$

where $Z^{(2d)}$ is a $2d$ -dimensional standard normal random vector. The data matrix whose columns consist of n i.i.d. samples from the distribution 14 is written as,

$$\begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix} = \begin{bmatrix} \Sigma_X^{(d)} & \Sigma_{XY}^{(d)} \\ \left(\Sigma_{XY}^{(d)}\right)^T & \Sigma_Y^{(d)} \end{bmatrix} \mathbf{Z}^{(2d)}, \quad (15)$$

where the columns of $\mathbf{Z}^{(2d)}$ consist of n i.i.d. samples from $2d$ -dimensional standard normal distribution. Denote by z_1 and z_2 the first and $(d+1)$ th rows of $\mathbf{Z}^{(2d)}$ corresponding to the first rows of $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$ respectively. Then, as $d \rightarrow \infty$ with the sample size n being fixed, the limiting behaviors of the sample canonical correlation coefficient $\hat{\rho}_i^{(d)}$ and its corresponding sample canonical weight vectors $\hat{\psi}_{X_i}^{(d)}$ and $\hat{\psi}_{Y_i}^{(d)}$ obtained from the data 15 are as follows,

Theorem 1 (Main result of the HDLSS asymptotic analysis of CCA.). (i) $\alpha > 1$

$$\begin{aligned} & \text{angle} \left(\hat{\psi}_{X1}^{(d)}, \psi_X^{(d)} \right) \xrightarrow{P}_{d \rightarrow \infty} \theta_X, \quad \text{angle} \left(\hat{\psi}_{Y1}^{(d)}, \psi_Y^{(d)} \right) \xrightarrow{P}_{d \rightarrow \infty} \theta_Y, \quad \hat{\rho}_1^{(d)} \xrightarrow{P}_{d \rightarrow \infty} \frac{\langle m_1, m_2 \rangle}{\|m_1\|_2 \|m_2\|_2}, \\ & \text{angle} \left(\hat{\psi}_{Xi}^{(d)}, \psi_X^{(d)} \right) \xrightarrow{P}_{d \rightarrow \infty} 0, \quad \text{angle} \left(\hat{\psi}_{Yi}^{(d)}, \psi_Y^{(d)} \right) \xrightarrow{P}_{d \rightarrow \infty} 0, \quad \hat{\rho}_i^{(d)} \xrightarrow{P}_{d \rightarrow \infty} 0, \quad i = 2, 3, \dots, n, \end{aligned}$$

where

$$\begin{aligned} m_1 &= (\sqrt{C_1}A_1^2 + \sqrt{C_2}B_1^2)z_1 + (\sqrt{C_1}A_1A_2 + \sqrt{C_2}B_1B_2)z_2, \\ m_2 &= (\sqrt{C_1}A_1A_2 + \sqrt{C_2}B_1B_2)z_1 + (\sqrt{C_1}A_1^2 + \sqrt{C_2}B_1^2)z_2, \end{aligned}$$

where

$$z_1, z_2 \stackrel{i.i.d.}{\sim} N \left(\begin{bmatrix} 0 \\ \mathbf{I} \end{bmatrix}, \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix} \right),$$

$$\begin{aligned}
C_1 &= \frac{\sigma_X^2 + \sigma_Y^2 + \sqrt{(\sigma_X^2)^2 - 2\sigma_X^2\sigma_Y^2 + 4\sigma_X^2\sigma_Y^2\rho^2 + (\sigma_Y^2)^2}}{2}, \\
C_2 &= \frac{\sigma_X^2 + \sigma_Y^2 - \sqrt{(\sigma_X^2)^2 - 2\sigma_X^2\sigma_Y^2 + 4\sigma_X^2\sigma_Y^2\rho^2 + (\sigma_Y^2)^2}}{2}, \\
A_1 &= \frac{C_1 - \sigma_Y^2}{\rho\sigma_X\sigma_Y} / \sqrt{\left(\frac{C_1 - \sigma_Y^2}{\rho\sigma_X\sigma_Y}\right)^2 + 1}, \quad A_2 = 1 / \sqrt{\left(\frac{C_1 - \sigma_Y^2}{\rho\sigma_X\sigma_Y}\right)^2 + 1}, \\
B_1 &= \frac{C_2 - \sigma_Y^2}{\rho\sigma_X\sigma_Y} / \sqrt{\left(\frac{C_2 - \sigma_Y^2}{\rho\sigma_X\sigma_Y}\right)^2 + 1}, \quad B_2 = 1 / \sqrt{\left(\frac{C_2 - \sigma_Y^2}{\rho\sigma_X\sigma_Y}\right)^2 + 1}.
\end{aligned}$$

(ii) $\alpha < 1$

$$\text{angle}\left(\hat{\psi}_{Xi}^{(d)}, \psi_X^{(d)}\right) \xrightarrow[d \rightarrow \infty]{P} 0, \quad \text{angle}\left(\hat{\psi}_{Yi}^{(d)}, \psi_Y^{(d)}\right) \xrightarrow[d \rightarrow \infty]{P} 0, \quad \hat{\rho}_i^{(d)} \xrightarrow[d \rightarrow \infty]{P} 1, \quad i = 1, 2, \dots, n.$$

3.2 Interpretation

The theorem 1 implies that where $\hat{\psi}_{X1}^{(d)}$ and $\hat{\psi}_{Y1}^{(d)}$ converge to depend heavily on the size of the variance d^α of the population eigenvector $\xi_{X1}^{(d)}$ and $\xi_{Y1}^{(d)}$. That is, the estimates $\hat{\psi}_{X1}^{(d)}$ and $\hat{\psi}_{Y1}^{(d)}$ tend to converge to the eigenvectors $\xi_{X1}^{(d)}$ and $\xi_{Y1}^{(d)}$ when their eigenvalues $\sigma_X^2 d^\alpha$ and $\sigma_Y^2 d^\alpha$ become strong enough ($\alpha > 1$) as $d \rightarrow \infty$. Briefly, we summarize results. The sample canonical weight vector $\hat{\psi}_{X1}^{(d)}$ (similarly $\hat{\psi}_{Y1}^{(d)}$) is,

- Consistent with $\psi_X^{(d)}$ if $\alpha > 1$ and $\text{angle}(\psi_X^{(d)}, \xi_{X1}^{(d)}) = 0$ as $d \rightarrow \infty$,
- Inconsistent with $\psi_X^{(d)}$ if $\alpha > 1$ and $\text{angle}(\psi_X^{(d)}, \xi_{X1}^{(d)}) = \theta_X$, for $0 < \theta_X < \pi/2$, as $d \rightarrow \infty$,
- Strongly inconsistent with $\psi_X^{(d)}$ if $\alpha < 1$ or if $\alpha > 1$ and $\text{angle}(\psi_X^{(d)}, \xi_{X1}^{(d)}) = \pi/2$ as $d \rightarrow \infty$.

The asymptotic behavior of the sample canonical correlation coefficient $\hat{\rho}_1^{(d)}$ is not straightforward to imagine. Let's take a simple example where $\sigma_X^2 = 1, \sigma_X^2 = 1, \tau_X^2 = 1$ and $\tau_Y^2 = 1$ in the spiked covariance structure in (5) and (6). In this case, referring to the theorem 1, the sample canonical correlation coefficient $\hat{\rho}_1^{(d)}$ converges in probability to the following random quantity,

$$\hat{\rho}_1^{(d)} \xrightarrow[d \rightarrow \infty]{P} \frac{\langle m_1, m_2 \rangle}{\|m_1\|_2 \|m_2\|_2},$$

where

$$\begin{aligned} m_1 &= \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) z_1 + \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) z_2, \\ m_2 &= \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) z_1 + \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) z_2. \end{aligned}$$

Note that z_1 and z_2 are samples from n -dimensional multivariate standard normal distribution. It can be easily verified that each element m_{1i} of m_1 (similarly for m_{2i} of m_2) follows a standard normal distribution,

$$\begin{aligned} m_{1,i} &= \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) z_{1i} + \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) z_{2i} \sim N(0, 1), \\ m_{2,i} &= \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) z_{1i} + \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) z_{2i} \sim N(0, 1), \end{aligned}$$

which leads to,

$$\|m_1\|_2 \sim \sqrt{\chi_n^2}, \quad \|m_2\|_2 \sim \sqrt{\chi_n^2},$$

where χ_n^2 denotes the chi-square distribution with degree of freedom of n . Since the numerator part $\langle m_1, m_2 \rangle$ is not a degenerate random quantity, one sees that $\hat{\rho}_1^{(d)}$ does not converge to a trivial random variable such as 1.

Now increase the sample size n to see which value the sample canonical correlation coefficient $\hat{\rho}_1^{(d)}$ converges to. By the law of large numbers and noting that the elements $m_{1,i}$ and $m_{2,i}$ are from i.i.d. standard normal distribution,

$$\frac{\|m_1\|_2^2}{n} = \sum_{i=1}^n \frac{m_{1i}^2}{n} \xrightarrow[n \rightarrow \infty]{P} 1, \quad \frac{\|m_2\|_2^2}{n} = \sum_{j=1}^n \frac{m_{2j}^2}{n} \xrightarrow[n \rightarrow \infty]{P} 1.$$

Furthurmore, noting that m_1 and m_2 are i.i.d. samples,

$$\begin{aligned} \frac{\langle m_1, m_2 \rangle}{n} &= \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) \sum_{i=1}^n \frac{z_{1i}^2}{n} \\ &\quad + \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) \sum_{i=1}^n \frac{z_{2i}^2}{n} \end{aligned}$$

$$\begin{aligned}
& + \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right)^2 \sum_{i=1}^n \frac{z_{1i} z_{2i}}{n} \\
& + \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right)^2 \sum_{i=1}^n \frac{z_{1i} z_{2i}}{n} \\
& \xrightarrow[n \rightarrow \infty]{P} 2 \left(\frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} \right) \left(\frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} \right) = \rho,
\end{aligned}$$

which confirms the conventional large sample asymptotic property of the statistic $\hat{\rho}_1^{(d)}$,

$$\hat{\rho}_1^{(d)} \xrightarrow[d, n \rightarrow \infty]{P} \rho.$$

4 Proof

We state here the theorem on the HDLSS asymptotic behavior of the sample eigenvalues and vectors (or sample PC variances and directions) included in [4] as frequently referred in this chapter.

Theorem 2 (HDLSS asymptotic result of engenvalues and engenvectors.). *Under the Gaussian assumption and the one spike case of (5) and (6),*

(i) *the limit of the first sample eigenvalue $\hat{\lambda}_{X1}^{(d)}$ (similarly for $\hat{\lambda}_{Yi}^{(d)}$) described in (12) depends on α ,*

$$\frac{\hat{\lambda}_{X1}^{(d)}}{\max(d^\alpha, d)} \Rightarrow \begin{cases} \sigma_X^2 \frac{\chi_n^2}{n}, & \alpha > 1, \\ \sigma_X^2 \frac{\chi_n^2}{n} + \frac{\tau_X^2}{n}, & \alpha = 1, \\ \frac{\tau_X^2}{n}, & \alpha < 1, \end{cases}$$

as $d \rightarrow \infty$, where \Rightarrow denotes the convergence in distribution, and χ_n^2 denotes the chi-square distribution with degree of freedom n . The rest of eigenvalues converge to the same quantities when scaled, that is, for any $\alpha \in [0, \infty)$, $i = 2, 3, \dots, n$,

$$\frac{\hat{\lambda}_{Xi}^{(d)}}{n} \rightarrow \frac{\tau_X^2}{n}, \text{ as } d \rightarrow \infty,$$

in probability.

(ii) The limit of the first eigenvectors $\hat{\xi}_{Xi}^{(d)}$ (similarly for $\hat{\xi}_{Yi}^{(d)}$) described in (12) depends on α ,

$$\langle \hat{\xi}_{Xi}^{(d)}, \xi_{Xi}^{(d)} \rangle \Rightarrow \begin{cases} 1, & \alpha > 1, \\ \left(1 + \frac{\tau_X^2}{\sigma_X^2 \chi_n^2}\right)^{-\frac{1}{2}}, & \alpha = 1, \\ 0, & \alpha < 1, \end{cases}$$

as $d \rightarrow \infty$. The rest of the eigenvectors are strongly inconsistent with their population counterpart, for any $\alpha \in [0, \infty)$, $i = 2, 3, \dots, n$,

$$\langle \hat{\xi}_{Xi}^{(d)}, \xi_{Xi}^{(d)} \rangle \rightarrow 0, \text{ as } \infty,$$

in probability.

4.1 Settings

Let $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$ be $d \times n$ matrices that collect d -dimensional vector observations measured on the common set of n samples. Following the representation (11) of the two random vectors $X^{(d)}$ and $Y^{(d)}$, we, using covariance and cross-covariance structures given in (8), (9) and (10), write $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$ as,

$$\begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix} = \left(\boldsymbol{\Sigma}_T^{(2d)} \right)^{\frac{1}{2}} \mathbf{Z}^{(2d)} = \begin{bmatrix} \boldsymbol{\Sigma}_X^{(d)} & \boldsymbol{\Sigma}_{XY}^{(d)} \\ \left(\boldsymbol{\Sigma}_{XY}^{(d)} \right)^T & \boldsymbol{\Sigma}_Y^{(d)} \end{bmatrix}^{\frac{1}{2}} \mathbf{Z}^{(2d)}, \quad (16)$$

where $\mathbf{Z}^{(2d)}$ is a $2d \times n$ matrix with columns of $\mathbf{Z}^{(2d)}$ being i.i.d. observations from an $2d$ -dimensional standard normal distribution. We introduce notations for elements of the matrix $\mathbf{Z}^{(2d)}$ for a later

use in the proof,

$$\mathbf{Z}^{(2d)} = \begin{bmatrix} \mathbf{Z}_X^{(d)} \\ \mathbf{Z}_Y^{(d)} \end{bmatrix} = \begin{bmatrix} z_{X11} & z_{X12} & \dots & z_{X1n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{Xd1} & z_{Xd2} & \dots & z_{Xdn} \\ z_{Y11} & z_{Y12} & \dots & z_{Y1n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{Yd1} & z_{Yd2} & \dots & z_{Ydn} \end{bmatrix} = \begin{bmatrix} z_{X1\bullet} \\ \vdots \\ z_{Xd\bullet} \\ z_{Y1\bullet} \\ \vdots \\ z_{Yd\bullet} \end{bmatrix}, \quad (17)$$

where $\mathbf{Z}_X^{(d)}$ denotes the upper half of $\mathbf{Z}^{(2d)}$ correspondint to $\mathbf{X}^{(d)}$ (similarly for $\mathbf{Z}_Y^{(d)}$), z_{Xij} , for $i = 1, 2, \dots, d$ and $j = 1, 2, \dots, n$, represents the i th element of the j observation in $\mathbf{Z}_X^{(d)}$ (similarly for z_{Yij}), and $z_{Xk\bullet}$ denotes the k th row of $\mathbf{Z}^{(2d)}$ (similarly for $z_{Yk\bullet}$). To investigate the asymptotic behavior of the sample covariance and cross-covariance of $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$, we want to expand the matrix in (16) to get an explicit expression elementwise. This can be done either by manual or using symbolic operations in Matlab. The results, however, is too long to be included in this page, so we are going to work with a big O, small o representation for the elements that have lengthy expressions. The covariance matrix $\Sigma_T^{(2d)}$ takes the following eigendecomposition,

$$\Sigma_T^{(2d)} = \sum_{i=1}^{2d} \lambda_{Ti}^{(2d)} \xi_{Ti}^{(2d)} \left(\xi_{Ti}^{(2d)} \right)^T, \quad (18)$$

where

$$\begin{aligned} \lambda_{T1}^{(2d)} &= O_P(d^\alpha), \lambda_{T2}^{(2d)} = O_P(d^\alpha), \lambda_{T4}^{(2d)} = O_P(1), \lambda_{T4}^{(2d)} = O_P(1), \\ \lambda_{Ti}^{(2d)} &= \tau_X^2, \quad i = 5, 6, \dots, d+2 \\ \lambda_{Tj}^{(2d)} &= \tau_Y^2, \quad j = d+3, d+4, \dots, 2d, \\ \xi_{T1}^{(2d)} &= \left[O_P(1) \quad O_P\left(\frac{1}{\sqrt{d^\alpha}}\right) \quad O_P(1) \quad \dots \quad O_P(1) \quad O_P(1) \quad O_P\left(\frac{1}{\sqrt{d^\alpha}}\right) \quad O_P(1) \quad \dots \quad O_P(1) \right]^T, \\ \xi_{T2}^{(2d)} &= \left[O_P(1) \quad O_P\left(\frac{1}{\sqrt{d^\alpha}}\right) \quad O_P(1) \quad \dots \quad O_P(1) \quad O_P(1) \quad O_P\left(\frac{1}{\sqrt{d^\alpha}}\right) \quad O_P(1) \quad \dots \quad O_P(1) \right]^T, \\ \xi_{T3}^{(2d)} &= \left[O_P(1) \quad O_P(1) \quad \dots \quad O_P(1) \quad O_P(1) \right]^T, \end{aligned}$$

$$\xi_{T4}^{(2d)} = \begin{bmatrix} O_P(1) & O_P(1) & \dots & O_P(1) & O_P(1) \end{bmatrix}^T,$$

$$\xi_{Ti}^{(2d)} = e_i^{(2d)}, \quad i = 5, 6, \dots, d+2,$$

$$\xi_{Tj}^{(2d)} = e_j^{(2d)}, \quad j = d+3, d+4, \dots, 2d.$$

Then, using (16), (17) and (18), the data matrix $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$ can be expressed as,

$$\begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix} = \begin{bmatrix} a_1 z_{X1\bullet} + a_2 z_{X2\bullet} + a_3 z_{Y1\bullet} + a_4 z_{Y2\bullet} \\ a_5 z_{X1\bullet} + a_6 z_{X2\bullet} + a_7 z_{Y1\bullet} + a_8(1) z_{Y2\bullet} \\ \tau_X z_{X3\bullet} \\ \vdots \\ \tau_X z_{Xd\bullet} \\ b_1 z_{X1\bullet} + b_2 z_{X2\bullet} + b_3 z_{Y1\bullet} + b_4 z_{Y2\bullet} \\ b_5 z_{X1\bullet} + b_6 z_{X2\bullet} + b_7 z_{Y1\bullet} + b_8 z_{Y2\bullet} \\ \tau_Y z_{Y3\bullet} \\ \vdots \\ \tau_Y z_{Yd\bullet} \end{bmatrix}, \quad (19)$$

where a_1, a_3, b_1 and b_3 are random variable of magnitude of $O_P(\sqrt{d^\alpha})$, and the rest of a_i and b_j are of magnitude of $O_P(1)$. Let c_1, c_2, d_1 and d_2 denote the first, second, $(d+1)$ th and $(d+2)$ th row of the matrix (19), respectively.

$$\begin{aligned} c_1 &= a_1 z_{X1\bullet} + a_2 z_{X2\bullet} + a_3 z_{Y1\bullet} + a_4 z_{Y2\bullet}, & c_2 &= a_5 z_{X1\bullet} + a_6 z_{X2\bullet} + a_7 z_{Y1\bullet} + a_8 z_{Y2\bullet}, \\ d_1 &= b_1 z_{X1\bullet} + b_2 z_{X2\bullet} + b_3 z_{Y1\bullet} + b_4 z_{Y2\bullet}, & d_2 &= b_5 z_{X1\bullet} + b_6 z_{X2\bullet} + b_7 z_{Y1\bullet} + b_8 z_{Y2\bullet}. \end{aligned} \quad (20)$$

The sample covariance and cross-covariance matrices $\hat{\Sigma}_X^{(d)}$, $\hat{\Sigma}_Y^{(d)}$ and $\hat{\Sigma}_{XY}^{(d)}$ are found as blocks of the following matrix,

$$\begin{bmatrix} n\hat{\Sigma}_X^{(d)} & n\hat{\Sigma}_{XY}^{(d)} \\ \left(n\hat{\Sigma}_{XY}^{(d)}\right)^T & n\hat{\Sigma}_Y^{(d)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix} \begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix}^T, \quad (21)$$

where elementwise-explicit matrices are given as,

$$\hat{\Sigma}_X^{(d)} = \frac{1}{n} \begin{bmatrix} \langle c_1, c_1 \rangle & \langle c_1, c_2 \rangle & \tau_X \langle c_1, z_{X3\bullet} \rangle & \tau_X \langle c_1, z_{X4\bullet} \rangle & \dots & \tau_X \langle c_1, z_{Xd\bullet} \rangle \\ \langle c_2, c_1 \rangle & \langle c_2, c_2 \rangle & \tau_X \langle c_2, z_{X3\bullet} \rangle & \tau_X \langle c_2, z_{X4\bullet} \rangle & \dots & \tau_X \langle c_2, z_{Xd\bullet} \rangle \\ \tau_X \langle z_{X3\bullet}, c_1 \rangle & \tau_X \langle z_{X3\bullet}, c_2 \rangle & \tau_X^2 \langle z_{X3\bullet}, z_{X3\bullet} \rangle & \tau_X^2 \langle z_{X3\bullet}, z_{X4\bullet} \rangle & \dots & \tau_X^2 \langle z_{X3\bullet}, z_{Xd\bullet} \rangle \\ \tau_X \langle z_{X4\bullet}, c_1 \rangle & \tau_X \langle z_{X4\bullet}, c_2 \rangle & \tau_X^2 \langle z_{X4\bullet}, z_{X3\bullet} \rangle & \tau_X^2 \langle z_{X4\bullet}, z_{X4\bullet} \rangle & \dots & \tau_X^2 \langle z_{X4\bullet}, z_{Xd\bullet} \rangle \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tau_X \langle z_{Xd\bullet}, c_1 \rangle & \tau_X \langle z_{Xd\bullet}, c_2 \rangle & \tau_X^2 \langle z_{Xd\bullet}, z_{X3\bullet} \rangle & \tau_X^2 \langle z_{Xd\bullet}, z_{X4\bullet} \rangle & \dots & \tau_X^2 \langle z_{Xd\bullet}, z_{Xd\bullet} \rangle \end{bmatrix},$$

$$\hat{\Sigma}_Y^{(d)} = \frac{1}{n} \begin{bmatrix} \langle d_1, d_1 \rangle & \langle d_1, d_2 \rangle & \tau_Y \langle d_1, z_{Y3\bullet} \rangle & \tau_Y \langle d_1, z_{Y4\bullet} \rangle & \dots & \tau_Y \langle d_1, z_{Yd\bullet} \rangle \\ \langle d_2, d_1 \rangle & \langle d_2, d_2 \rangle & \tau_Y \langle d_2, z_{Y3\bullet} \rangle & \tau_Y \langle d_2, z_{Y4\bullet} \rangle & \dots & \tau_Y \langle d_2, z_{Yd\bullet} \rangle \\ \tau_Y \langle z_{Y3\bullet}, d_1 \rangle & \tau_Y \langle z_{Y3\bullet}, d_2 \rangle & \tau_Y^2 \langle z_{Y3\bullet}, z_{Y3\bullet} \rangle & \tau_Y^2 \langle z_{Y3\bullet}, z_{Y4\bullet} \rangle & \dots & \tau_Y^2 \langle z_{Y3\bullet}, z_{Yd\bullet} \rangle \\ \tau_Y \langle z_{Y4\bullet}, d_1 \rangle & \tau_Y \langle z_{Y4\bullet}, d_2 \rangle & \tau_Y^2 \langle z_{Y4\bullet}, z_{Y3\bullet} \rangle & \tau_Y^2 \langle z_{Y4\bullet}, z_{Y4\bullet} \rangle & \dots & \tau_Y^2 \langle z_{Y4\bullet}, z_{Yd\bullet} \rangle \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tau_Y \langle z_{Yd\bullet}, d_1 \rangle & \tau_Y \langle z_{Yd\bullet}, d_2 \rangle & \tau_Y^2 \langle z_{Yd\bullet}, z_{Y3\bullet} \rangle & \tau_Y^2 \langle z_{Yd\bullet}, z_{Y4\bullet} \rangle & \dots & \tau_Y^2 \langle z_{Yd\bullet}, z_{Yd\bullet} \rangle \end{bmatrix}, \quad (22)$$

$$\hat{\Sigma}_{XY}^{(d)} = \frac{1}{n} \begin{bmatrix} \langle c_1, d_1 \rangle & \langle c_1, d_2 \rangle & \tau_Y \langle c_1, z_{Y3\bullet} \rangle & \tau_Y \langle c_1, z_{Y4\bullet} \rangle & \dots & \tau_Y \langle c_1, z_{Yd\bullet} \rangle \\ \langle c_2, d_1 \rangle & \langle c_2, d_2 \rangle & \tau_Y \langle c_2, z_{Y3\bullet} \rangle & \tau_Y \langle c_2, z_{Y4\bullet} \rangle & \dots & \tau_Y \langle c_2, z_{Yd\bullet} \rangle \\ \tau_X \langle z_{X3\bullet}, d_1 \rangle & \tau_X \langle z_{X3\bullet}, d_2 \rangle & \tau_X \tau_Y \langle z_{X3\bullet}, z_{Y3\bullet} \rangle & \tau_X \tau_Y \langle z_{X3\bullet}, z_{Y4\bullet} \rangle & \dots & \tau_X \tau_Y \langle z_{X3\bullet}, z_{Yd\bullet} \rangle \\ \tau_X \langle z_{X4\bullet}, d_1 \rangle & \tau_X \langle z_{X4\bullet}, d_2 \rangle & \tau_X \tau_Y \langle z_{X4\bullet}, z_{Y3\bullet} \rangle & \tau_X \tau_Y \langle z_{X4\bullet}, z_{Y4\bullet} \rangle & \dots & \tau_X \tau_Y \langle z_{X4\bullet}, z_{Yd\bullet} \rangle \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tau_X \langle z_{Xd\bullet}, d_1 \rangle & \tau_X \langle z_{Xd\bullet}, d_2 \rangle & \tau_X \tau_Y \langle z_{Xd\bullet}, z_{Y3\bullet} \rangle & \tau_X \tau_Y \langle z_{Xd\bullet}, z_{Y4\bullet} \rangle & \dots & \tau_X \tau_Y \langle z_{Xd\bullet}, z_{Yd\bullet} \rangle \end{bmatrix}.$$

4.2 Notation

Here, we introduce notations to be used in the proof for HDLSS asymptotic behaviors of a random variable.

- For a d -dimensional random variable $X^{(d)}$, we say that $X^{(d)} = o_P(d^\alpha)$, for $\alpha \in R$, if, $\forall \epsilon > 0$,

$$\lim_{d \rightarrow \infty} P \left(\left| \frac{X^{(d)}}{d^\alpha} > \epsilon \right| \right) = 0.$$

- For a d -dimensional random variable $X^{(d)}$, we say that $X^{(d)} = O_P(d^\alpha)$, for $\alpha \in R$, if, $\forall \epsilon > 0$, there exists a finite $C > 0$ such that,

$$P \left(\left| \frac{X^{(d)}}{d^\alpha} > M \right| \right) < \epsilon, \quad \forall d.$$

- For a d -dimensional random variable $X^{(d)}$, we say that $X^{(d)} \asymp d^\alpha$, for $\alpha \in R$, if,

$$X^{(d)} = O_P(d^\alpha) \text{ and } X^{(d)} \neq o_P(d^\alpha).$$

4.3 Case of $\alpha > 1$

First, we prove Theorem 1 under the condition of $\alpha > 1$ with the spiked model of the population covariance structure of $X^{(d)}$ and $Y^{(d)}$ described in (9).

4.3.1 Behavior of the sample cross-covariance matrix

We now investigate the HDLSS asymptotic behavior of the sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$ given in (22), in specific, its sample singular values and singular vectors. The singular value decomposition (SVD) of $\hat{\Sigma}_{XY}^{(d)}$ gives,

$$\hat{\Sigma}_{XY}^{(d)} = \sum_{i=1}^n \hat{\lambda}_{XYi}^{(d)} \hat{\eta}_{Xi}^{(d)} \left(\hat{\eta}_{Yi}^{(d)} \right)^T, \quad (23)$$

where $\hat{\lambda}_{XY1}^{(d)} \geq \hat{\lambda}_{XY2}^{(d)} \geq \dots \geq \hat{\lambda}_{XYn}^{(d)} \geq 0$, $\|\hat{\eta}_{Xi}^{(d)}\|_2 = \|\hat{\eta}_{Yi}^{(d)}\|_2 = 1$ for $i = 1, 2, \dots, n$, and $\langle \hat{\eta}_{Xi}^{(d)}, \hat{\eta}_{Xj}^{(d)} \rangle = \langle \hat{\eta}_{Yi}^{(d)}, \hat{\eta}_{Yj}^{(d)} \rangle = 0$, for $i \neq j$.

Lemma 1 (CCA HDLSS Asymptotic lemma 1.). *Let C_{XY} and \mathbf{M}_{XY} be,*

$$C_{XY} = \lim_{d \rightarrow \infty} \frac{\langle c_1, d_1 \rangle}{d^\alpha}, \quad \mathbf{M}_{XY} = \begin{bmatrix} C_{XY} & \mathbf{0}_{1 \times (d-1)} \\ \mathbf{0}_{(d-1) \times 1} & \mathbf{0}_{(d-1) \times (d-1)} \end{bmatrix},$$

where c_1 and d_1 are defined in (20) and so C_{XY} is a non-degenerate random variable. Then, for $\alpha > 1$,

$$\left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \mathbf{M}_{XY} \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0,$$

where $\|\bullet\|_F$ is the Frobenious norm of a matrix.

Proof. Let $\mathbf{M}_{XY}^{(d)}$ be,

$$\mathbf{M}_{XY}^{(d)} = \begin{bmatrix} \frac{\langle c_1, d_1 \rangle}{d^\alpha} & \mathbf{0}_{1 \times (d-1)} \\ \mathbf{0}_{(d-1) \times 1} & \mathbf{0}_{(d-1) \times (d-1)} \end{bmatrix}.$$

It is obvious to see that,

$$\left\| \mathbf{M}_{XY}^{(d)} - \mathbf{M}_{XY} \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Using the Cauchy-Schwarz inequality,

$$\begin{aligned} \left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \mathbf{M}_{XY}^{(d)} \right\|_F^2 &= \frac{\langle c_1, d_2 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, d_1 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, d_2 \rangle^2}{d^{2\alpha}} + \tau_Y^2 \sum_{i=3}^d \frac{\langle c_1, z_{Yi\bullet} \rangle^2}{d^{2\alpha}} \\ &\quad + \tau_Y^2 \sum_{i=3}^d \frac{\langle c_2, z_{Yi\bullet} \rangle^2}{d^{2\alpha}} + \tau_X^2 \sum_{i=3}^d \frac{\langle z_{Xi\bullet}, d_1 \rangle^2}{d^{2\alpha}} \\ &\quad + \tau_X^2 \sum_{i=3}^d \frac{\langle z_{Xi\bullet}, d_2 \rangle^2}{d^{2\alpha}} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\langle z_{Xi\bullet}, z_{Yj\bullet} \rangle^2}{d^{2\alpha}} \\ &\leq \frac{\langle c_1, d_2 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, d_1 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, d_2 \rangle^2}{d^{2\alpha}} + \tau_Y^2 \sum_{i=3}^d \frac{\|c_1\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^{2\alpha}} \\ &\quad + \tau_Y^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^{2\alpha}} + \tau_X^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|d_1\|_2^2}{d^{2\alpha}} \end{aligned}$$

$$+ \tau_X^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|d_2\|_2^2}{d^{2\alpha}} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Yj\bullet}\|_2^2}{d^{2\alpha}}.$$

Since $\langle c_1, d_2 \rangle^2$, $\langle c_2, d_1 \rangle^2$ are $O_P(\sqrt{d^\alpha})$, and $\langle c_2, d_2 \rangle^2$ is $O_P(1)$,

$$\frac{\langle c_1, d_2 \rangle^2}{d^{2\alpha}} \xrightarrow{p} 0, \quad \frac{\langle c_2, d_1 \rangle^2}{d^{2\alpha}} \xrightarrow{p} 0, \quad \frac{\langle c_2, d_2 \rangle^2}{d^{2\alpha}} \xrightarrow{p} 0.$$

It is not hard to see that $\|z_{Xi\bullet}\|_2^2 \sim \chi_n^2$ and $\|z_{Yi\bullet}\|_2^2 \sim \chi_n^2$, since the elements of $z_{Xi\bullet}$ (respectively $z_{Yi\bullet}$) are the i th components of the i.i.d. multivariate standard normal samples of size n . Note that the magnitudes of $\|c_1\|_2^2$, $\|c_2\|_2^2$, $\|d_1\|_2^2$ and $\|d_2\|_2^2$ are of $O_P(d^\alpha)$. Applying the law of large numbers, we have,

$$\begin{aligned} \tau_Y^2 \sum_{i=3}^d \frac{\|c_1\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_Y^2}{d^{\alpha-1}} \left\| \frac{c_1}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Yi\bullet}\|_2^2}{d} \\ &\xrightarrow{p} 0 \times O_P(1) \times E(\chi_n^2) = 0, \\ \tau_Y^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_Y^2}{d^{\alpha-1}} \left\| \frac{c_2}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Yi\bullet}\|_2^2}{d} \\ &\xrightarrow{p} 0 \times 0 \times E(\chi_n^2) = 0, \\ \tau_X^2 \sum_{i=3}^d \frac{\|d_1\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_X^2}{d^{\alpha-1}} \left\| \frac{d_1}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \\ &\xrightarrow{p} 0 \times O_P(1) \times E(\chi_n^2) = 0, \\ \tau_X^2 \sum_{i=3}^d \frac{\|d_2\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_X^2}{d^{\alpha-1}} \left\| \frac{d_2}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \\ &\xrightarrow{p} 0 \times 0 \times E(\chi_n^2) = 0. \end{aligned}$$

Using $\alpha > 1$ and the law of large numbers,

$$\begin{aligned} \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Yj\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_X^2 \tau_Y^2}{d^{2\alpha-2}} \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \sum_{j=3}^d \frac{\|z_{Yj\bullet}\|_2^2}{d} \\ &\xrightarrow{p} 0 \times E(\chi_n^2) \times E(\chi_n^2) = 0. \end{aligned}$$

Therefore,

$$\left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \mathbf{M}_{XY}^{(d)} \right\|_F^2 \xrightarrow{d \rightarrow \infty} 0.$$

□

Lemma 2 (CCA HDLSS Asymptotic lemma 2.). *Let \mathbf{M}_{XY} be the matrix defined in Lemma 1. Let $\hat{\lambda}_{XY1}^{(d)}$, $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ be the first sample singular value and singular vectors from (23). Then, for $\alpha > 1$,*

$$\left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)} \right)^T - \mathbf{M}_{XY} \right\|_F^2 \xrightarrow{d \rightarrow \infty} 0.$$

Proof. Using $\mathbf{M}_{XY}^{(d)}$ defined in Lemma 1 and the triangle inequality,

$$\begin{aligned} & \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)} \right)^T - \mathbf{M}_{XY}^{(d)} \right\|_F \\ &= \left\| \left(\frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \mathbf{M}_{XY}^{(d)} \right) - \left(\frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)} \right)^T \right) \right\|_F \\ &\leq \left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \mathbf{M}_{XY}^{(d)} \right\|_F + \left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)} \right)^T \right\|_F. \end{aligned}$$

By Lemma 1,

$$\left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \mathbf{M}_{XY}^{(d)} \right\|_F^2 \xrightarrow{d \rightarrow \infty} 0.$$

Write $\mathbf{M}_{XY}^{(d)}$ as,

$$\mathbf{M}_{XY}^{(d)} = \frac{\langle c_1, d_1 \rangle}{d^\alpha} e_1^{(d)} \left(e_1^{(d)} \right)^T.$$

Since the first sample singular value $\hat{\lambda}_{XY1}^{(d)}$ and singular vectors $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ provide the best rank-1

approximation to $\hat{\lambda}_{XY1}^{(d)}/d^\alpha$,

$$\begin{aligned} \left\| \frac{\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)} \right)^T \right\|_F^2 &\leq \left\| \frac{\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \frac{\langle c_1, d_1 \rangle}{nd^\alpha} e_1^{(d)} \left(e_1^{(d)} \right)^T \right\|_F^2 \\ &= \left\| \frac{\hat{\Sigma}_{XY}^{(d)}}{d^\alpha} - \frac{\mathbf{M}_{XY}^{(d)}}{n} \right\|_F^2 \\ &\xrightarrow[d \rightarrow \infty]{p} 0. \end{aligned}$$

□

Lemma 3 (CCA HDLSS Asymptotic lemma 3.). *For $\alpha > 1$, the first singular value $\hat{\lambda}_{XY1}^{(d)}$ and singular vectors $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ converge in probability to the following quantities as $d \rightarrow \infty$,*

$$\frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \xrightarrow[d \rightarrow \infty]{p} C_{XY}, \quad \left\| \hat{\eta}_{X1}^{(d)} - e_1^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 0, \quad \left\| \hat{\eta}_{Y1}^{(d)} - e_1^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 0,$$

where C_{XY} is defined in Lemma 1.

Proof. Let $\tilde{\eta}_{X1}^{(d)}$ and $\tilde{\eta}_{Y1}^{(d)}$ be $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ where their first entries are set to 0. Let $\hat{\eta}_{X1}^{(d)}(i)$ and $\hat{\eta}_{Y1}^{(d)}(j)$ be the i th and j th entries of $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$. By Lemma 2,

$$\begin{aligned} \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{Y1}^{(d)} \right\|_2^2 &= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=2}^d \left(\hat{\eta}_{X1}^{(d)}(i) \right)^2 \sum_{j=2}^d \left(\hat{\eta}_{Y1}^{(d)}(j) \right)^2 \\ &= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=2}^d \sum_{j=2}^d \left(\hat{\eta}_{X1}^{(d)}(i) \right)^2 \left(\hat{\eta}_{Y1}^{(d)}(j) \right)^2 \\ &\xrightarrow[d \rightarrow \infty]{p} 0. \end{aligned}$$

First, we show that $(\hat{\lambda}_{XY1}^{(d)}/d^\alpha)^2 > 0$ in probability. Suppose that $(\hat{\lambda}_{XY1}^{(d)}/d^\alpha)^2$ converges in probability to 0. Then, noting that $\left\| \hat{\eta}_{X1}^{(d)} \right\|_2^2 = 1$ and $\left\| \hat{\eta}_{Y1}^{(d)} \right\|_2^2 = 1$,

$$\begin{aligned} \left\| \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)} \right)^T \right\|_F^2 &= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=1}^d \sum_{j=1}^d \left(\hat{\eta}_{X1}^{(d)}(i) \right)^2 \left(\hat{\eta}_{Y1}^{(d)}(j) \right)^2 \\ &= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=1}^d \left(\hat{\eta}_{X1}^{(d)}(i) \right)^2 \sum_{j=1}^d \left(\hat{\eta}_{Y1}^{(d)}(j) \right)^2 \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \left\| \hat{\eta}_{X1}^{(d)} \right\|_2^2 \left\| \hat{\eta}_{Y1}^{(d)} \right\|_2^2 \\
&= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \\
&\xrightarrow[d \rightarrow \infty]{p} 0,
\end{aligned}$$

which contradicts to Lemma 2 (note that the limiting matrix M is not a degenerate matrix).

Therefore,

$$\left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 0, \text{ or } \left\| \tilde{\eta}_{Y1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 0.$$

We want to show that both $\|\tilde{\eta}_{X1}^{(d)}\|_2^2$ and $\|\tilde{\eta}_{Y1}^{(d)}\|_2^2$ converge to 0 in probability. Suppose that $\|\tilde{\eta}_{X1}^{(d)}\|_2^2$ converges to 0 and $\|\tilde{\eta}_{Y1}^{(d)}\|_2^2 > 0$ in probability. Since the norm of $\hat{\eta}_{X1}^{(d)}$ is 1, its first entry $\hat{\eta}_{X1}^{(d)}(1)$ converges in probability to,

$$\left(\hat{\eta}_{X1}^{(d)}(1) \right)^2 = 1 - \left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 1.$$

Then, the squared sum of the entries in the first row of $(\hat{\lambda}_{XY1}^{(d)}/d^\alpha)\hat{\eta}_{X1}^{(d)}\left(\hat{\eta}_{Y1}^{(d)}\right)^T$ with the first entry excluded becomes,

$$\left\| \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)}(1) \tilde{\eta}_{Y1}^{(d)} \right\|_2^2 = \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \left(\hat{\eta}_{X1}^{(d)}(1) \right)^2 \sum_{j=2}^d \left(\hat{\eta}_{Y1}^{(d)}(j) \right)^2 \xrightarrow[d \rightarrow \infty]{p} > 0,$$

which contradicts to Lemma 2, according to which the above quantity should converge to 0. Argument is similar for the case where $\|\tilde{\eta}_{Y1}^{(d)}\|_2^2$ converges to 0 and $\|\tilde{\eta}_{X1}^{(d)}\|_2^2 > 0$ in probability. Hence we have

$$\left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 0, \text{ and } \left\| \tilde{\eta}_{Y1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 0.$$

Note that, since the norm of $\hat{\eta}_{Y1}^{(d)}$ is 1,

$$\left(\hat{\eta}_{Y1}^{(d)}(1) \right)^2 = 1 - \left\| \tilde{\eta}_{Y1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{p} 1.$$

Therefore

$$\begin{aligned}\left\|\hat{\eta}_{X1}^{(d)} - e_1^{(d)}\right\|_2^2 &= \left(\hat{\eta}_{X1}^{(d)}(1) - 1\right)^2 + \left\|\tilde{\eta}_{X1}^{(d)}(i)\right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 0, \\ \left\|\hat{\eta}_{Y1}^{(d)} - e_1^{(d)}\right\|_2^2 &= \left(\hat{\eta}_{Y1}^{(d)}(1) - 1\right)^2 + \left\|\tilde{\eta}_{Y1}^{(d)}(i)\right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 0.\end{aligned}$$

To calculate the limiting value of $\hat{\lambda}_{XY1}^{(d)}/d^\alpha$ as $d \rightarrow \infty$, using the unitary invariance property of the Frobenius norm and the previous result about the limiting vectors of $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ as $d \rightarrow \infty$,

$$\begin{aligned}& \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)}\right)^T - \mathbf{M}_{XY}^{(d)} \right\|_F^2 \\&= \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)}\right)^T - \frac{\langle c_1, d_1 \rangle}{d^\alpha} e_1^{(d)} \left(e_1^{(d)}\right)^T \right\|_F^2 \\&= \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \left(\hat{\eta}_{Y1}^{(d)}\right)^T \hat{\eta}_{X1}^{(d)} \left(\hat{\eta}_{Y1}^{(d)}\right)^T \hat{\eta}_{X1}^{(d)} - \frac{\langle c_1, d_1 \rangle}{d^\alpha} \left(\hat{\eta}_{Y1}^{(d)}\right)^T e_1^{(d)} \left(e_1^{(d)}\right)^T \hat{\eta}_{X1}^{(d)} \right\|_F^2 \\&= \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} - \frac{\langle c_1, d_1 \rangle}{d^\alpha} \left(\hat{\eta}_{Y1}^{(d)}\right)^T e_1^{(d)} \left(e_1^{(d)}\right)^T \hat{\eta}_{X1}^{(d)} \right\|_F^2 \\&= \left(\frac{\langle c_1, d_1 \rangle}{d^\alpha} \right)^2 \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}/d^\alpha}{\langle c_1, d_1 \rangle/d^\alpha} - \left(\hat{\eta}_{Y1}^{(d)}\right)^T e_1^{(d)} \left(e_1^{(d)}\right)^T \hat{\eta}_{X1}^{(d)} \right\|_F^2 \\&\xrightarrow[p]{d \rightarrow \infty} A^2 \left(\frac{B}{C} - 1 \right)^2.\end{aligned}$$

where,

$$A = \lim_{d \rightarrow \infty} \frac{\langle c_1, d_1 \rangle}{d^\alpha}, \quad B = \lim_{d \rightarrow \infty} \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha}, \quad C = \lim_{d \rightarrow \infty} \frac{\langle c_1, d_1 \rangle}{d^\alpha}.$$

By Lemma 2,

$$\left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \hat{\eta}_{X1}^{(d)} \hat{\eta}_{Y1}^{(d)} - \mathbf{M}_{XY}^{(d)} \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Since $A \asymp O_p(1)$,

$$\left(\frac{B}{C} - 1 \right)^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Therefore,

$$\frac{B}{C} \xrightarrow[p]{d \rightarrow \infty} 1.$$

□

Lemma 4 (CCA HDLSS Asymptotic lemma 4.). *The magnitudes of the entries $\hat{\eta}_{X1}^{(d)}(i)$ and $\hat{\eta}_{Y1}^{(d)}(i)$, $i = 2, 3, \dots, d$, of the first sample singular vectors $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ are of $O_p(1/\sqrt{d^\alpha})$ as $d \rightarrow \infty$.*

Proof. Let $\hat{\Sigma}_{XY}^{(d)}(i, j)$ be the entry of the i th row and j th column of $\hat{\Sigma}_{XY}^{(d)}$. Consider $\hat{\Sigma}_{XY}^{(d)}(1, 3)$, which is $O_p(\sqrt{d^\alpha})$. The contribution of $\hat{\lambda}_{XY1}^{(d)}$, $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{Y1}^{(d)}$ to $\hat{\Sigma}_{XY}^{(d)}(1, 3)$ is,

$$\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1) \hat{\eta}_{Y1}^{(d)}(3).$$

From Lemma 3,

$$\hat{\lambda}_{XY1}^{(d)} \asymp O_p(d^\alpha), \quad \hat{\eta}_{X1}^{(d)}(1) \xrightarrow[p]{d \rightarrow \infty} 1.$$

Therefore,

$$\hat{\eta}_{Y1}^{(d)}(3) = O_p\left(\frac{1}{\sqrt{d^\alpha}}\right).$$

Argument is similar for the rest of entries. □

Lemma 5 (CCA HDLSS Asymptotic lemma 5.). *The magnitudes of the sample singular values $\hat{\lambda}_{XYi}^{(d)}$ for $i = 2, 3, \dots, n$ are of $O_p(d)$.*

Proof. Let $\tilde{\Sigma}_{XY}^{(d)}$ be the $(d-1) \times (d-1)$ sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$ from which the first row and column are set to 0. Let $\tilde{\eta}_{Xi}^{(d)}$ and $\tilde{\eta}_{Yi}^{(d)}$ be $\hat{\eta}_{Xi}^{(d)}$ and $\hat{\eta}_{Yi}^{(d)}$ where their first entries are set to 0. Then,

$$\begin{aligned} \frac{n\hat{\lambda}_{XY1}^{(d)}}{d} \tilde{\eta}_{X1}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)}\right)^T + \sum_{i=2}^n \frac{n\hat{\lambda}_{XYi}^{(d)}}{d} \tilde{\eta}_{Xi}^{(d)} \left(\tilde{\eta}_{Yi}^{(d)}\right)^T &= \frac{n\tilde{\Sigma}_{XY}^{(d)}}{d}, \\ \sum_{i=2}^n \frac{n\hat{\lambda}_{XYi}^{(d)}}{d} \tilde{\eta}_{Xi}^{(d)} \left(\tilde{\eta}_{Yi}^{(d)}\right)^T &= \frac{n\tilde{\Sigma}_{XY}^{(d)}}{d} - \frac{n\hat{\lambda}_{XY1}^{(d)}}{d} \tilde{\eta}_{X1}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)}\right)^T, \end{aligned}$$

$$\left\| \sum_{i=2}^n \frac{n\hat{\lambda}_{XYi}^{(d)}}{d} \tilde{\eta}_{Xi}^{(d)} \left(\tilde{\eta}_{Yi}^{(d)} \right)^T \right\|_F \leq \left\| \frac{n\tilde{\Sigma}_{XY}^{(d)}}{d} \right\|_F + \left\| \frac{n\hat{\lambda}_{XY1}^{(d)}}{d} \tilde{\eta}_{X1}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)} \right)^T \right\|_F.$$

Note that $\|c_2\|_2^2$ and $\|d_2\|_2^2$ are $O_P(1)$ and that $\|z_{Yi\bullet}\|_2^2 \sim \chi_n^2$ and $\|z_{Xi\bullet}\|_2^2 \sim \chi_n^2$. Using the Cauchy-Schwarz inequality,

$$\begin{aligned} \left\| \frac{n\tilde{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 &= \frac{\langle c_2, d_2 \rangle^2}{d^2} + \tau_Y^2 \sum_{i=3}^d \frac{\langle c_2, z_{Yi\bullet} \rangle^2}{d^2} \\ &\quad + \tau_X^2 \sum_{i=3}^d \frac{\langle z_{Xi\bullet}, d_2 \rangle^2}{d^2} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\langle z_{Xi\bullet}, z_{Yj\bullet} \rangle^2}{d^2} \\ &\leq \frac{\langle c_2, d_2 \rangle^2}{d^2} + \tau_Y^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^2} \\ &\quad + \tau_X^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|d_2\|_2^2}{d^2} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Yj\bullet}\|_2^2}{d^2} \\ &= \frac{\langle c_2, d_2 \rangle^2}{d^2} + \tau_Y^2 \left\| \frac{c_2}{\sqrt{d}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Yi\bullet}\|_2^2}{d} \\ &\quad + \tau_Y^2 \left\| \frac{d_2}{\sqrt{d}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Yi\bullet}\|_2^2}{d} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \sum_{j=3}^d \frac{\|z_{Yj\bullet}\|_2^2}{d} \\ &\xrightarrow{d \rightarrow \infty} 0 + 0 + 0 + \tau_X^2 \tau_Y^2 E^2(\chi_n^2) \\ &= \tau_X^2 \tau_Y^2 n^2. \end{aligned}$$

Using Lemma 3 and 4 on the magnitudes of $\hat{\lambda}_{XY1}^{(d)}$ and the entries of $\hat{\eta}_{X1}^{(d)}(i)$ and $\hat{\eta}_{Y1}^{(d)}(i)$,

$$\hat{\lambda}_{XY1}^{(d)} \asymp O_p(d^\alpha), \quad \hat{\eta}_{X1}^{(d)}(i) = O_p\left(\frac{1}{\sqrt{d^\alpha}}\right), i = 2, 3, \dots, d.$$

Then we have,

$$\begin{aligned} \left\| \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \tilde{\eta}_{X1}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)} \right)^T \right\|_F^2 &= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \sum_{i=2}^d \sum_{j=2}^d \left(\hat{\eta}_{X1}^{(d)}(i) \right)^2 \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 \\ &= \left(\hat{\lambda}_{XY1}^{(d)} \right)^2 \sum_{i=2}^d \frac{\left(\hat{\eta}_{X1}^{(d)}(i) \right)^2}{d} \sum_{j=2}^d \frac{\left(\hat{\eta}_{X1}^{(d)}(j) \right)^2}{d} \\ &= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=2}^d \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(i) \right)^2}{d} \sum_{j=2}^d \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(j) \right)^2}{d} \end{aligned}$$

$$= O_p(1).$$

Hence

$$\left\| \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \tilde{\eta}_{Xi}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)} \right)^T \right\|_F^2 = O_p(1). \quad (24)$$

Expanding the squared Frobenius norm of the above matrix,

$$\begin{aligned} \left\| \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \tilde{\eta}_{Xi}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)} \right)^T \right\|_F^2 &= \sum_{j=2}^d \sum_{k=2}^d \left(\sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(j) \hat{\eta}_{Yi}^{(d)}(k) \right)^2 \\ &= \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \sum_{j=2}^d \sum_{k=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \right)^2 \left(\hat{\eta}_{Yi}^{(d)}(k) \right)^2 \\ &\quad + 2 \sum_{i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \sum_{j=2}^d \sum_{k=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \hat{\eta}_{Xi'}^{(d)}(j) \right) \left(\hat{\eta}_{Yi}^{(d)}(k) \hat{\eta}_{Yi'}^{(d)}(k) \right) \\ &= \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \sum_{j=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \right)^2 \sum_{k=2}^d \left(\hat{\eta}_{Yi}^{(d)}(k) \right)^2 \\ &\quad + 2 \sum_{i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \sum_{j=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \hat{\eta}_{Xi'}^{(d)}(j) \right) \sum_{k=2}^d \left(\hat{\eta}_{Yi}^{(d)}(k) \hat{\eta}_{Yi'}^{(d)}(k) \right) \\ &= \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{Yi}^{(d)} \right\|_2^2 \\ &\quad + 2 \sum_{2 \leq i, i' \leq n, i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \left\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \right\rangle \left\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \right\rangle. \end{aligned}$$

We, to simplify the last part in the above equalities, show that $\|\tilde{\eta}_{Xi}^{(d)}\|_2^2$ and $\|\tilde{\eta}_{Yi}^{(d)}\|_2^2$, for $i = 2, 3, \dots, n$, approach to 1 and that $\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \rangle$ and $\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \rangle$, for $i, i' = 2, 3, \dots, n$ and $i \neq i'$, approach to 0 as $d \rightarrow \infty$. Consider $\|\tilde{\eta}_{X2}^{(d)}\|_2^2$. Since $\|\hat{\eta}_{X1}^{(d)} - e_1^{(d)}\|_2$ converges in probability to 0 by Lemma 3 and $\|\hat{\eta}_{X1}^{(d)}\|_2^2 = 1$,

$$\left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 = 1 - \left(\hat{\eta}_{X1}^{(d)}(1) \right)^2 \xrightarrow{d \rightarrow \infty} 0,$$

which leads to,

$$\left\langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{X2}^{(d)} \right\rangle^2 \leq \left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{X2}^{(d)} \right\|_2^2 \leq \left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \times 1 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Using the orthogonality of $\hat{\eta}_{X1}^{(d)}$ and $\hat{\eta}_{X2}^{(d)}$,

$$\left\langle \hat{\eta}_{X1}^{(d)}, \hat{\eta}_{X2}^{(d)} \right\rangle = \hat{\eta}_{X1}^{(d)}(1) \hat{\eta}_{X2}^{(d)}(1) + \left\langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{X2}^{(d)} \right\rangle = 0.$$

Since $\hat{\eta}_{X1}^{(d)}(1)$ converges in probability to 1 by Lemma 3, we have,

$$\hat{\eta}_{X2}^{(d)}(1) \xrightarrow[p]{d \rightarrow \infty} 0.$$

Since $\left\| \hat{\eta}_{X2}^{(d)} \right\|_2^2 = 1$,

$$\left\| \tilde{\eta}_{X2}^{(d)} \right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 1.$$

Simiarly,

$$\left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 1, \quad \left\| \tilde{\eta}_{Yi}^{(d)} \right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 1, \quad i = 2, 3, \dots, n.$$

Now consider $\langle \tilde{\eta}_{X2}^{(d)}, \tilde{\eta}_{X3}^{(d)} \rangle$. Since $\hat{\eta}_{X2}^{(d)}$ and $\hat{\eta}_{X3}^{(d)}$ are orthogonal,

$$\left\langle \hat{\eta}_{X2}^{(d)}, \hat{\eta}_{X3}^{(d)} \right\rangle = \hat{\eta}_{X2}^{(d)}(1) \hat{\eta}_{X3}^{(d)}(1) + \left\langle \tilde{\eta}_{X2}^{(d)}, \tilde{\eta}_{X3}^{(d)} \right\rangle = 0.$$

We showed that $\left\| \tilde{\eta}_{X2}^{(d)} \right\|_2^2$ and $\left\| \tilde{\eta}_{X3}^{(d)} \right\|_2^2$ converge in probability to 1, which implies,

$$\left\langle \tilde{\eta}_{X2}^{(d)}, \tilde{\eta}_{X3}^{(d)} \right\rangle \xrightarrow[p]{d \rightarrow \infty} 0.$$

Simiarly,

$$\left\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \right\rangle \xrightarrow[p]{d \rightarrow \infty} 0, \quad \left\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \right\rangle \xrightarrow[p]{d \rightarrow \infty} 0, \quad i, i' = 2, 3, \dots, n, \quad i \neq i'.$$

Then, by (24),

$$\begin{aligned}
& \left\| \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \tilde{\eta}_{Xi}^{(d)} \left(\tilde{\eta}_{Y1}^{(d)} \right)^T \right\|_F^2 \\
&= \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{Yi}^{(d)} \right\|_2^2 + 2 \sum_{2 \leq i, i' \leq n, i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \left\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \right\rangle \left\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \right\rangle \\
&\xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 = O_p(1).
\end{aligned}$$

Hence, the magnitudes of $\hat{\lambda}_{XYi}^{(d)}$, $i = 2, 3, \dots, n$ are of $O_p(d)$ as $d \rightarrow \infty$. \square

Lemma 6 (CCA HDLSS Asymptotic lemma 6.). *This lemma improves Lemma 5 by providing a precise limiting value. The sample singular values $\hat{\lambda}_{XYi}^{(d)}$ scaled by $1/d$, for $i = 2, 3, \dots, n$, converge in probability to the following quantity as $d \rightarrow \infty$,*

$$\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} 0, \quad i = 2, 3, \dots, n.$$

Proof. Let $\tilde{\Sigma}_{XY}^{(d)}$ be the $(d-1) \times (d-1)$ sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$ from which the first row and column are set to 0. We showed that,

$$\left\| \frac{\tilde{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 \xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} \leq \tau_X^2 \tau_Y^2. \quad (25)$$

Let $\tilde{\eta}_{Xi}^{(d)}$ and $\tilde{\eta}_{Yi}^{(d)}$ be $\hat{\eta}_{Xi}^{(d)}$ and $\hat{\eta}_{Yi}^{(d)}$ where their first entries are set to 0. Then,

$$\begin{aligned}
\left\| \frac{\tilde{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 &= \sum_{j=2}^d \sum_{k=2}^d \left(\sum_{i=1}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(j) \hat{\eta}_{Yi}^{(d)}(k) \right)^2 \\
&= \sum_{i=1}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \sum_{j=2}^d \sum_{k=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \right)^2 \left(\hat{\eta}_{Yi}^{(d)}(k) \right)^2 \\
&\quad + 2 \sum_{1 \leq i, i' \leq n, i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \sum_{j=2}^d \sum_{k=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \hat{\eta}_{Xi'}^{(d)}(j) \right) \left(\hat{\eta}_{Yi}^{(d)}(k) \hat{\eta}_{Yi'}^{(d)}(k) \right) \\
&= \sum_{i=1}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \sum_{j=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \right)^2 \sum_{k=2}^d \left(\hat{\eta}_{Yi}^{(d)}(k) \right)^2
\end{aligned}$$

$$\begin{aligned}
& + 2 \sum_{1 \leq i, i' \leq n, i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \sum_{j=2}^d \left(\hat{\eta}_{Xi}^{(d)}(j) \hat{\eta}_{Xi'}^{(d)}(j) \right) \sum_{k=2}^d \left(\hat{\eta}_{Yi}^{(d)}(k) \hat{\eta}_{Yi'}^{(d)}(k) \right) \\
& = \frac{\left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \sum_{j=2}^d \left(\tilde{\eta}_{X1}^{(d)}(j) \right)^2 \sum_{k=2}^d \left(\tilde{\eta}_{Y1}^{(d)}(k) \right)^2}{(1)} \\
& + 2 \sum_{i=2}^n \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left\langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{Xi}^{(d)} \right\rangle \left\langle \tilde{\eta}_{Y1}^{(d)}, \tilde{\eta}_{Yi}^{(d)} \right\rangle + \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{Yi}^{(d)} \right\|_2^2 \\
& \quad \quad \quad (2) \quad \quad \quad (3) \\
& + 2 \sum_{2 \leq i, i' \leq n, i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \left\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \right\rangle \left\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \right\rangle. \\
& \quad \quad \quad (4)
\end{aligned}$$

In the proof of Lemma 5, we showed that $\|\tilde{\eta}_{Xi}^{(d)}\|_2^2$ and $\|\tilde{\eta}_{Yi}^{(d)}\|_2^2$, for $i = 2, 3, \dots, n$, converge in probability to 1 and that $\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \rangle$ and $\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \rangle$, for $i, i' = 2, 3, \dots, n$ and $i \neq i'$, converge in probability to 1 as $d \rightarrow \infty$. Since $\hat{\lambda}_{XYi}$, for $i = 2, 3, \dots, n$, is of magnitude of $O_p(d)$ by Lemma 5, we have for (3) and (4),

$$\begin{aligned}
& \left| \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{Yi}^{(d)} \right\|_2^2 - \sum_{i=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \right| \xrightarrow[d \rightarrow \infty]{p} 0, \\
& 2 \sum_{2 \leq i, i' \leq n, i \neq i'} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \frac{\hat{\lambda}_{XYi'}^{(d)}}{d} \left\langle \tilde{\eta}_{Xi}^{(d)}, \tilde{\eta}_{Xi'}^{(d)} \right\rangle \left\langle \tilde{\eta}_{Yi}^{(d)}, \tilde{\eta}_{Yi'}^{(d)} \right\rangle \xrightarrow[d \rightarrow \infty]{p} 0.
\end{aligned}$$

Since $\|\tilde{\eta}_{Xi}^{(d)}\|_2^2$ and $\|\tilde{\eta}_{Yi}^{(d)}\|_2^2$, for $i = 2, 3, \dots, n$, converge in probability to 1, the probability that an infinite number of entries of $\tilde{\eta}_{Xi}^{(d)}$ and $\tilde{\eta}_{Yi}^{(d)}$ are of $(\asymp 1/d^a)$, for $0 \leq a < 1/2$, is 0. Suppose that an infinite number of entries of $\tilde{\eta}_{Xi}^{(d)}$ are of $(\asymp 1/d^a)$, for $0 \leq a < 1/2$. Then, the squared sum of its entries blows up,

$$\sum_{j=2}^d \left(\tilde{\eta}_{Xi}^{(d)}(j) \right)^2 = \frac{1}{d^{2a-1}} \sum_{j=2}^d \frac{d^{2a} \tilde{\eta}_{Xi}^{(d)}(j)}{d} \xrightarrow[d \rightarrow \infty]{p} \infty, \quad 0 \leq a < \frac{1}{2}.$$

Hence, only a finite number of entries of $\tilde{\eta}_{Xi}^{(d)}$ and $\tilde{\eta}_{Yi}^{(d)}$ are of magnitude of $(\asymp 1/d^a)$, for $0 \leq a < 1/2$

and the rest of entries are of $O_P(1/\sqrt{d})$ as $d \rightarrow \infty$. Using this fact and Lemma 4,

$$\begin{aligned}
\langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{Xi}^{(d)} \rangle &= \sum_{j=2}^d \tilde{\eta}_{X1}^{(d)}(j) \tilde{\eta}_{Xi}^{(d)}(j) \\
&= \sum_{j \in I} \tilde{\eta}_{X1}^{(d)}(j) \tilde{\eta}_{Xi}^{(d)}(j) + \sum_{j \notin I} \tilde{\eta}_{X1}^{(d)}(j) \tilde{\eta}_{Xi}^{(d)}(j) \\
&= \sum_{j \in I} \tilde{\eta}_{X1}^{(d)}(j) \tilde{\eta}_{Xi}^{(d)}(j) + \frac{1}{\sqrt{d^{\alpha-1}}} \sum_{j \notin I} \frac{\left(\sqrt{d^\alpha} \tilde{\eta}_{X1}^{(d)}(j) \right) \left(\sqrt{d} \tilde{\eta}_{Xi}^{(d)}(j) \right)}{d} \\
&= O_P(1/\sqrt{d^{\alpha-1}}), \quad i = 2, 3, \dots, n,
\end{aligned}$$

where I is an index set denoting the entries of $\tilde{\eta}_{Xi}^{(d)}$ of magnitude of $(\asymp 1/d^a)$, for $0 \leq a < 1/2$. Similarly, the magnitude of $\langle \tilde{\eta}_{Y1}^{(d)}, \tilde{\eta}_{Yi}^{(d)} \rangle$ is $O_P(1/\sqrt{d^{\alpha-1}})$, for $i = 2, 3, \dots, n$. Hence, by Lemma 3 and 5,

$$\begin{aligned}
(2) &= 2 \sum_{i=2}^n \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{Xi}^{(d)} \rangle \langle \tilde{\eta}_{Y1}^{(d)}, \tilde{\eta}_{Yi}^{(d)} \rangle \\
&= 2 \sum_{i=2}^n \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{Xi}^{(d)} \rangle \right) \left(\sqrt{d^{\alpha-1}} \langle \tilde{\eta}_{Y1}^{(d)}, \tilde{\eta}_{Yi}^{(d)} \rangle \right) \\
&= O_P(1).
\end{aligned}$$

We prove that the term (2) above indeed converges to 0 as $d \rightarrow \infty$. With Lemma 4 and the fact that $\langle \hat{\eta}_{X1}^{(d)}, \hat{\eta}_{Xi}^{(d)} \rangle = 0$ and $\|\tilde{\eta}_{Xi}^{(d)}\|$ converges in probability to 0, for $i = 2, 3, \dots, n$, the Cauchy-Schwarz inequality implies,

$$\begin{aligned}
\left(\hat{\eta}_{X1}^{(d)}(1) \hat{\eta}_{Xi}^{(d)}(1) \right)^2 &= \left(- \sum_{j=2}^d \hat{\eta}_{X1}^{(d)}(j) \hat{\eta}_{Xi}^{(d)}(j) \right)^2 \\
&= \langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{Xi}^{(d)} \rangle^2 \\
&\leq \left\| \tilde{\eta}_{X1}^{(d)} \right\|_2^2 \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \\
&= \sum_{j=2}^d \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2 \\
&= \frac{1}{d^{\alpha-1}} \sum_{j=2}^d \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(j) \right)^2}{d} \left\| \tilde{\eta}_{Xi}^{(d)} \right\|_2^2
\end{aligned}$$

$$= O_P\left(\frac{1}{d^{\alpha-1}}\right).$$

Since $\hat{\eta}_{X1}^{(d)}(1)$ converges in probability to 1 by Lemma 3, we have,

$$\hat{\eta}_{Xi}^{(d)}(1) = O_P\left(\frac{1}{\sqrt{d^{\alpha-1}}}\right), \quad i = 2, 3, \dots, n. \quad (26)$$

Similarly,

$$\hat{\eta}_{Yi}^{(d)}(1) = O_P\left(\frac{1}{\sqrt{d^{\alpha-1}}}\right), \quad i = 2, 3, \dots, n.$$

Consider $\hat{\Sigma}_{XY}^{(d)}(3, 1)$, an entry in the third row and first column of the sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$ given in (22). Then, we have,

$$\begin{aligned} \frac{\hat{\Sigma}_{XY}^{(d)}(3, 1)}{\sqrt{d^\alpha}} &= \frac{\tau_X \langle z_{X3\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} \\ &= \frac{1}{\sqrt{d^\alpha}} \sum_{i=1}^n \hat{\lambda}_{XYi}^{(d)} \hat{\eta}_{Xi}^{(d)}(3) \hat{\eta}_{Yi}^{(d)}(1) \\ &= \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(3) \right) \hat{\eta}_{Y1}^{(d)}(1) + \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{\sqrt{d^{2\alpha-1}}} \hat{\eta}_{Xi}^{(d)}(3) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right), \end{aligned}$$

which implies that,

$$\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(3) = \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \left(\frac{\tau_X \langle z_{X3\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} - \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{\sqrt{d^{2\alpha-1}}} \hat{\eta}_{Xi}^{(d)}(3) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right) \right).$$

In a similar argument as above, it can be shown that,

$$\begin{aligned} \sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(j) &= \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} - \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{\sqrt{d^{2\alpha-1}}} \hat{\eta}_{Xi}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right) \right), \quad j = 4, 5, \dots, d, \\ \sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(k) &= \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1)} \left(\frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n\sqrt{d^\alpha}} - \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{\sqrt{d^{2\alpha-1}}} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xi}^{(d)}(1) \right) \hat{\eta}_{Yi}^{(d)}(k) \right), \quad k = 3, 4, \dots, d. \end{aligned}$$

Then, the term (2) goes as follows,

$$(2) = 2 \sum_{i=2}^n \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left\langle \tilde{\eta}_{X1}^{(d)}, \tilde{\eta}_{Xi}^{(d)} \right\rangle \left\langle \tilde{\eta}_{Y1}^{(d)}, \tilde{\eta}_{Yi}^{(d)} \right\rangle$$

$$\begin{aligned}
&= 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=2}^d \hat{\eta}_{X1}^{(d)}(j) \hat{\eta}_{Xi}^{(d)}(j) \right) \left(\sum_{k=2}^d \hat{\eta}_{Y1}^{(d)}(k) \hat{\eta}_{Yi}^{(d)}(k) \right) \\
&= 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \hat{\eta}_{X1}^{(d)}(2) \hat{\eta}_{Y1}^{(d)}(2) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(2) \hat{\eta}_{Yi}^{(d)}(2) \\
&\quad + 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \hat{\eta}_{X1}^{(d)}(2) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(2) \left(\sum_{k=3}^d \hat{\eta}_{Y1}^{(d)}(k) \hat{\eta}_{Yi}^{(d)}(k) \right) \\
&\quad + 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \hat{\eta}_{Y1}^{(d)}(2) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Yi}^{(d)}(2) \left(\sum_{k=3}^d \hat{\eta}_{X1}^{(d)}(k) \hat{\eta}_{Xi}^{(d)}(k) \right) \\
&\quad + 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \hat{\eta}_{X1}^{(d)}(j) \hat{\eta}_{Xi}^{(d)}(j) \right) \left(\sum_{k=3}^d \hat{\eta}_{Y1}^{(d)}(k) \hat{\eta}_{Yi}^{(d)}(k) \right).
\end{aligned}$$

Using Lemma 3, 4 and the fact that only a finite number of entries of $\hat{\eta}_{Xi}^{(d)}$ and $\hat{\eta}_{Yi}^{(d)}$, for $i = 2, 3, \dots, n$, can be of magnitude $(\asymp 1/d^a)$, for $0 \leq a < 1/2$, the first three terms in the last part of the above equalities becomes,

$$\begin{aligned}
&2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \hat{\eta}_{X1}^{(d)}(2) \hat{\eta}_{Y1}^{(d)}(2) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(2) \hat{\eta}_{Yi}^{(d)}(2) \\
&= \frac{2}{d} \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(2) \right) \left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(2) \right) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(2) \hat{\eta}_{Yi}^{(d)}(2) \xrightarrow{d \rightarrow \infty} 0, \\
&2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \hat{\eta}_{X1}^{(d)}(2) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(2) \left(\sum_{k=3}^d \hat{\eta}_{Y1}^{(d)}(k) \hat{\eta}_{Yi}^{(d)}(k) \right) \\
&= \frac{2}{d} \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(2) \right) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d} \hat{\eta}_{Xi}^{(d)}(2) \right) \left(\sum_{k=3}^d \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(k) \right) \left(\sqrt{d} \hat{\eta}_{Yi}^{(d)}(k) \right)}{d} \right) \xrightarrow{d \rightarrow \infty} 0, \\
&2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \hat{\eta}_{Y1}^{(d)}(2) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Yi}^{(d)}(2) \left(\sum_{k=3}^d \hat{\eta}_{X1}^{(d)}(k) \hat{\eta}_{Xi}^{(d)}(k) \right) \\
&= \frac{2}{d} \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(2) \right) \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d} \hat{\eta}_{Yi}^{(d)}(2) \right) \left(\sum_{k=3}^d \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(k) \right) \left(\sqrt{d} \hat{\eta}_{Xi}^{(d)}(k) \right)}{d} \right) \xrightarrow{d \rightarrow \infty} 0.
\end{aligned}$$

The fourth term expands as,

$$2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \hat{\eta}_{X1}^{(d)}(j) \hat{\eta}_{Xi}^{(d)}(j) \right) \left(\sum_{k=3}^d \hat{\eta}_{Y1}^{(d)}(k) \hat{\eta}_{Yi}^{(d)}(k) \right)$$

$$\begin{aligned}
&= 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(j) \hat{\eta}_{Xi}^{(d)}(j) \right) \left(\sum_{k=3}^d \sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(k) \hat{\eta}_{Yi}^{(d)}(k) \right) \\
&= 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \\
&\quad \times \left(\sum_{j=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n \sqrt{d^\alpha}} - \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{\sqrt{d^{2\alpha-1}}} \hat{\eta}_{Xl}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yl}^{(d)}(1) \right) \right) \hat{\eta}_{Xi}^{(d)}(j) \right) \\
&\quad \times \left(\sum_{k=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1)} \left(\frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n \sqrt{d^\alpha}} - \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{\sqrt{d^{2\alpha-1}}} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xl}^{(d)}(1) \right) \hat{\eta}_{Yl}^{(d)}(k) \right) \hat{\eta}_{Yi}^{(d)}(k) \right).
\end{aligned}$$

Using the the Cauchy-Schwarz inequality on $\langle \hat{\eta}_{Xi}^{(d)}, \hat{\eta}_{Yj}^{(d)} \rangle$, for $i, j = 2, 3, \dots, n$, the law of large number and the fact that $\alpha > 1$ and only a finite number of entries of $\hat{\eta}_{Xi}^{(d)}$ and $\hat{\eta}_{Yi}^{(d)}$, for $i = 2, 3, \dots, n$, can be of magnitude ($\asymp 1/d^a$), for $0 \leq a < 1/2$, it is not hard to see that,

$$\begin{aligned}
&2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d} \sum_{i=2}^n \left[\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{\sqrt{d^{2\alpha-1}}} \hat{\eta}_{Xl}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yl}^{(d)}(1) \right) \right) \hat{\eta}_{Xi}^{(d)}(j) \right) \right. \\
&\quad \times \left. \left(\sum_{k=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1)} \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{\sqrt{d^{2\alpha-1}}} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xl}^{(d)}(1) \right) \hat{\eta}_{Yl}^{(d)}(k) \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \right] \\
&= 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d^{\alpha-1}} \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \right)^2 \sum_{i=2}^n \left[\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{d} \hat{\eta}_{Xl}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yl}^{(d)}(1) \right) \right) \hat{\eta}_{Xi}^{(d)}(j) \right) \right. \\
&\quad \times \left. \left(\sum_{k=3}^d \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xl}^{(d)}(1) \right) \hat{\eta}_{Yl}^{(d)}(k) \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \right] \xrightarrow{d \rightarrow \infty} 0, \\
&2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d} \sum_{i=2}^n \left[\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \sum_{j=3}^d \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n \sqrt{d^\alpha}} \hat{\eta}_{Xi}^{(d)}(j) \right) \right. \\
&\quad \times \left. \left(\sum_{k=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1)} \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{\sqrt{d^{2\alpha-1}}} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xl}^{(d)}(1) \right) \hat{\eta}_{Yl}^{(d)}(k) \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \right] \\
&= 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{\sqrt{d^{\alpha-1}}} \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \right)^2 \sum_{i=2}^n \left[\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \sum_{j=3}^d \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n \sqrt{d^\alpha}} \frac{\sqrt{d} \hat{\eta}_{Xi}^{(d)}(j)}{d} \right) \right. \\
&\quad \times \left. \left(\sum_{k=3}^d \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xl}^{(d)}(1) \right) \hat{\eta}_{Yl}^{(d)}(k) \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \right] \xrightarrow{d \rightarrow \infty} 0, \\
&2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d} \sum_{i=2}^n \left[\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \sum_{j=3}^d \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n \sqrt{d^\alpha}} \hat{\eta}_{Yi}^{(d)}(j) \right) \right.
\end{aligned}$$

$$\begin{aligned}
& \times \left(\sum_{k=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1)} \left(- \sum_{l=2}^n \frac{\hat{\lambda}_{XYl}^{(d)}}{\sqrt{d^{2\alpha-1}}} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xl}^{(d)}(1) \right) \hat{\eta}_{Yl}^{(d)}(k) \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \Big] \xrightarrow{d \rightarrow \infty} 0, \\
& 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \frac{1}{d} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n \sqrt{d^\alpha}} \right) \hat{\eta}_{Xi}^{(d)}(j) \right) \\
& \times \left(\sum_{k=3}^d \frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{X1}^{(d)}(1)} \left(\frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n \sqrt{d^\alpha}} \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \\
& = 2 \frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)} \hat{\eta}_{Y1}^{(d)}(1)} \right)^2 \sum_{i=2}^n \left[\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sum_{j=3}^d \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n \sqrt{d^\alpha}} \right) \frac{\sqrt{d} \hat{\eta}_{Xi}^{(d)}(j)}{d} \right) \right. \\
& \left. \times \left(\sum_{k=3}^d \left(\frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n \sqrt{d^\alpha}} \right) \frac{\sqrt{d} \hat{\eta}_{Yi}^{(d)}(k)}{d} \right) \right] \xrightarrow{d \rightarrow \infty} 0.
\end{aligned}$$

We now prove that the term (1) converges in probability to $\tau_X^2 \tau_Y^2$ as $d \rightarrow \infty$,

$$(1) = \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \sum_{j=2}^d \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 \sum_{k=2}^d \left(\hat{\eta}_{Y1}^{(d)}(k) \right)^2 \xrightarrow{d \rightarrow \infty} \tau_X^2 \tau_Y^2.$$

which implies that $\hat{\lambda}_{XYi}^{(d)}/d$, for $i = 2, 3, \dots, n$, are squeezed, by the condition of (25), to converge in probability to 0 as $d \rightarrow \infty$. The term (1) can be written as,

$$\begin{aligned}
& \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \sum_{j=2}^d \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 \sum_{k=2}^d \left(\hat{\eta}_{Y1}^{(d)}(k) \right)^2 \\
& = \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \left(\hat{\eta}_{X1}^{(d)}(2) \right)^2 \left(\hat{\eta}_{Y1}^{(d)}(2) \right)^2 + \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \left(\hat{\eta}_{X1}^{(d)}(2) \right)^2 \sum_{k=3}^d \left(\hat{\eta}_{Y1}^{(d)}(k) \right)^2 \\
& + \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \left(\hat{\eta}_{Y1}^{(d)}(2) \right)^2 \sum_{j=3}^d \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 + \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \sum_{j=3}^d \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 \sum_{k=3}^d \left(\hat{\eta}_{Y1}^{(d)}(k) \right)^2.
\end{aligned}$$

By the use of Lemma 3 and 4,

$$\begin{aligned}
& \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \left(\hat{\eta}_{X1}^{(d)}(2) \right)^2 \left(\hat{\eta}_{Y1}^{(d)}(2) \right)^2 = \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(2) \right)^2 \frac{1}{d^2} \left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(2) \right)^2 \xrightarrow{d \rightarrow \infty} 0, \\
& \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \left(\hat{\eta}_{X1}^{(d)}(2) \right)^2 \sum_{k=3}^d \left(\hat{\eta}_{Y1}^{(d)}(k) \right)^2 = \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(2) \right)^2}{d} \sum_{k=3}^n \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(k) \right)^2}{d}
\end{aligned}$$

$$\begin{aligned}
&= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \frac{1}{d} \frac{\sum_{k=3}^n \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(2) \right)^2 \left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(k) \right)^2}{d} \xrightarrow[p]{d \rightarrow \infty} 0, \\
&\left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \left(\hat{\eta}_{Y1}^{(d)}(2) \right)^2 \sum_{k=3}^d \left(\hat{\eta}_{X1}^{(d)}(k) \right)^2 = \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(2) \right)^2}{d} \sum_{k=3}^n \frac{\left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(k) \right)^2}{d} \\
&= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \frac{1}{d} \frac{\sum_{k=3}^n \left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(2) \right)^2 \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(k) \right)^2}{d} \xrightarrow[p]{d \rightarrow \infty} 0.
\end{aligned}$$

Note that the magnitudes of $\tau_X \langle z_{X3\bullet}, d_1 \rangle / n\sqrt{d^\alpha}$ and $\tau_Y \langle c_1, z_{Y3\bullet} \rangle / n\sqrt{d^\alpha}$ are of $(\asymp 1)$ from (20).

Using Lemma 3, 4 and (26),

$$\begin{aligned}
&\left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d} \right)^2 \sum_{j=3}^d \left(\hat{\eta}_{X1}^{(d)}(j) \right)^2 \sum_{k=3}^d \left(\hat{\eta}_{Y1}^{(d)}(k) \right)^2 \\
&= \left(\frac{\hat{\lambda}_{XY1}^{(d)}}{d^\alpha} \right)^2 \frac{1}{d^2} \sum_{j=3}^d \left(\sqrt{d^\alpha} \hat{\eta}_{X1}^{(d)}(j) \right)^2 \sum_{k=3}^d \left(\sqrt{d^\alpha} \hat{\eta}_{Y1}^{(d)}(k) \right)^2 \\
&= \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)}} \right)^2 \frac{1}{d^2} \sum_{j=3}^d \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} - \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{\sqrt{d^{2\alpha-1}}} \hat{\eta}_{Xi}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right) \right)^2 \\
&\quad \times \sum_{k=3}^d \left(\frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n\sqrt{d^\alpha}} - \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{\sqrt{d^{2\alpha-1}}} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xi}^{(d)}(1) \right) \hat{\eta}_{Yi}^{(d)}(k) \right)^2 \\
&= \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)}} \right)^2 \frac{1}{d^2} \sum_{j=3}^d \left(\frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} \right)^2 \sum_{k=3}^d \left(\frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n\sqrt{d^\alpha}} \right)^2 \\
&\quad - 4 \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)}} \right)^2 \frac{1}{d^{2\alpha-1}} \left(\sum_{j=3}^d \frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right) \right) \\
&\quad \times \left(\sum_{k=3}^d \frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n\sqrt{d^\alpha}} \sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xi}^{(d)}(1) \right) \hat{\eta}_{Yi}^{(d)}(k) \right) \\
&\quad + \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)}} \right)^2 \frac{1}{d^{4(\alpha-1)}} \sum_{j=3}^d \left(\sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \hat{\eta}_{Xi}^{(d)}(j) \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right) \right)^2 \\
&\quad \times \sum_{k=3}^d \left(\sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xi}^{(d)}(1) \right) \hat{\eta}_{Yi}^{(d)}(k) \right)^2.
\end{aligned}$$

The second term in the last equality above converges in probability to 0. Using the fact that $\alpha > 1$ and only a finite number of entries of $\hat{\eta}_{Xi}^{(d)}$ and $\hat{\eta}_{Yi}^{(d)}$, for $i = 2, 3, \dots, n$, can be of magnitude $(\asymp 1/d^\alpha)$,

for $0 \leq a < 1/2$,

$$\begin{aligned} & \frac{1}{d^{2\alpha-2}} \left(\sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right) \sum_{j=3}^d \frac{\tau_X \langle z_{Xj\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} \frac{\sqrt{d} \hat{\eta}_{Xi}^{(d)}(j)}{d} \right) \\ & \times \left(\sum_{i=2}^n \frac{\hat{\lambda}_{XYi}^{(d)}}{d} \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xi}^{(d)}(1) \right) \sum_{k=3}^d \frac{\tau_Y \langle c_1, z_{Yk\bullet} \rangle}{n\sqrt{d^\alpha}} \frac{\sqrt{d} \hat{\eta}_{Yi}^{(d)}(k)}{d} \right) \xrightarrow{d \rightarrow \infty} 0. \end{aligned}$$

The third term also converges in probability to 0. It easy to see the result with the following equivalent form, noting that each inner product is bounded by 1 by the Cauchy-Schwarz inequality and that $\alpha > 1$,

$$\begin{aligned} & \frac{1}{d^{4(\alpha-1)}} \left(\sum_{i,j=2}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \left(\frac{\hat{\lambda}_{XYj}^{(d)}}{d} \right)^2 \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Yi}^{(d)}(1) \right)^2 \left(\sqrt{d^{\alpha-1}} \hat{\eta}_{Xj}^{(d)}(1) \right)^2 \sum_{k=3}^d \left(\hat{\eta}_{Xi}^{(d)}(k) \right)^2 \left(\hat{\eta}_{Yj}^{(d)}(k) \right)^2 \right) \\ & + \frac{4}{d^{4(\alpha-1)}} \left(\sum_{2 \leq a, a' \leq n, a \neq a'} \sum_{2 \leq b, b' \leq n, b \neq b'} \frac{\hat{\lambda}_{XYa}^{(d)}}{d} \frac{\hat{\lambda}_{XYa'}^{(d)}}{d} \frac{\hat{\lambda}_{XYb}^{(d)}}{d} \frac{\hat{\lambda}_{XYb'}^{(d)}}{d} \sqrt{d^{\alpha-1}} \hat{\eta}_{Ya}^{(d)}(1) \sqrt{d^{\alpha-1}} \hat{\eta}_{Yb'}^{(d)}(1) \right. \\ & \quad \left. \times \sqrt{d^{\alpha-1}} \hat{\eta}_{Xb}^{(d)}(1) \sqrt{d^{\alpha-1}} \hat{\eta}_{Xb'}^{(d)}(1) \sum_{k=3}^d \left\langle \hat{\eta}_{Xa}^{(d)}(k), \hat{\eta}_{Xa'}^{(d)}(k) \right\rangle \left\langle \hat{\eta}_{Yb}^{(d)}(k), \hat{\eta}_{Yb'}^{(d)}(k) \right\rangle \right) \xrightarrow{d \rightarrow \infty} 0. \end{aligned}$$

Now, look at the first term. Denote by $d_1(i)$ and $c_1(i)$ the i th entries of the vectors d_1 and c_1 given in (20). Then, recalling that $z_{Xij}^2 \sim \chi_1^2$ (similarly z_{Yij}^2) and that z_{Xki} and z_{Xkj} are independent (similarly for z_{Yki} and z_{Ykj}) and using the limiting quantity of the first sample singular value in Lemma 3,

$$\begin{aligned} & \tau_X^2 \tau_Y^2 \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)}} \right)^2 \frac{1}{d^2} \sum_{j=3}^d \left(\frac{\langle z_{Xj\bullet}, d_1 \rangle}{n\sqrt{d^\alpha}} \right)^2 \sum_{k=3}^d \left(\frac{\langle c_1, z_{Yk\bullet} \rangle}{n\sqrt{d^\alpha}} \right)^2 \\ & = \tau_X^2 \tau_Y^2 \frac{1}{n^2} \left(\frac{d^\alpha}{\hat{\lambda}_{XY1}^{(d)}} \right)^2 \left(\sum_{i=1}^n \frac{d_1^2(i)}{d^\alpha} \sum_{j=3}^d \frac{z_{Xji}^2}{d} + 2 \sum_{1 \leq i, j \leq n, i \neq j} \sum_{k=1}^d \frac{d_1(i) d_1(j)}{d^\alpha} \frac{z_{Xki} z_{Xkj}}{d} \right) \\ & \quad \times \left(\sum_{i=1}^n \frac{c_1^2(i)}{d^\alpha} \sum_{j=3}^d \frac{z_{Yji}^2}{d} + 2 \sum_{1 \leq i, j \leq n, i \neq j} \sum_{k=1}^d \frac{c_1(i) c_1(j)}{d^\alpha} \frac{z_{Yki} z_{Ykj}}{d} \right) \\ & \xrightarrow{d \rightarrow \infty} \tau_X^2 \tau_Y^2 \frac{1}{n^2} \frac{n^2 \|c_1\|_2^2 \|d_1\|_2^2}{\langle c_1, d_1 \rangle^2} \\ & \geq \tau_X^2 \tau_Y^2, \end{aligned}$$

where the last inequality results from the Cauchy-Schwarz inequality of $\langle c_1, d_1 \rangle^2 \leq \|c_1\|_2^2 \|d_1\|_2^2$.

Then, the condition of (25) completes the proof. \square

4.3.2 Behavior of the sample covariance matrices

We investigate the HDLSS asymptotic behavior of the sample covariance matrices $\hat{\Sigma}_X^{(d)}$ and $\hat{\Sigma}_Y^{(d)}$ given in (22), in specific, its sample eigenvalues and eigenvectors. Here, we only include the result of the analysis of $\hat{\Sigma}_X^{(d)}$ as that of $\hat{\Sigma}_Y^{(d)}$ is similar. The eigendecomposition of $\hat{\Sigma}_X^{(d)}$ gives,

$$\hat{\Sigma}_X^{(d)} = \sum_{i=1}^n \hat{\lambda}_{Xi}^{(d)} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T, \quad (27)$$

where $\hat{\lambda}_{X1}^{(d)} \geq \hat{\lambda}_{X2}^{(d)} \geq \dots \geq \hat{\lambda}_{Xn}^{(d)} \geq 0$, $\|\hat{\xi}_{Xi}^{(d)}\|_2 = 1$ and $\langle \hat{\xi}_{Xi}^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle = 0$ for $i \neq j$.

Lemma 7 (CCA HDLSS Asymptotic lemma 7.). *Let C_X and \mathbf{M}_X be,*

$$C_X = \lim_{d \rightarrow \infty} \frac{\|c_1\|_2^2}{d^\alpha}, \quad \mathbf{M}_X = \begin{bmatrix} C_X & \mathbf{0}_{1 \times (d-1)} \\ \mathbf{0}_{(d-1) \times 1} & \mathbf{0}_{(d-1) \times (d-1)} \end{bmatrix}.$$

where c_1 is defined in (20) and so C_X is a non-degenerate random variable. Then, for $\alpha > 1$,

$$\left\| \frac{n \hat{\Sigma}_X^{(d)}}{d^\alpha} - \mathbf{M}_X \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0,$$

Proof. Let $\mathbf{M}_X^{(d)}$ be,

$$\mathbf{M}_X^{(d)} = \begin{bmatrix} \frac{\|c_1\|_2^2}{d^\alpha} & \mathbf{0}_{1 \times (d-1)} \\ \mathbf{0}_{(d-1) \times 1} & \mathbf{0}_{(d-1) \times (d-1)} \end{bmatrix}.$$

It is obvious to see that,

$$\left\| \mathbf{M}_X^{(d)} - \mathbf{M}_X \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left\| \frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \mathbf{M}_X^{(d)} \right\|_F^2 &= \frac{\langle c_1, c_2 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, c_1 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, c_2 \rangle^2}{d^{2\alpha}} + 2\tau_X^2 \sum_{i=3}^d \frac{\langle c_1, z_{Xi\bullet} \rangle^2}{d^{2\alpha}} \\
&\quad + 2\tau_X^2 \sum_{i=3}^d \frac{\langle c_2, z_{Xi\bullet} \rangle^2}{d^{2\alpha}} + \tau_X^4 \sum_{i=3}^d \sum_{j=3}^d \frac{\langle z_{Xi\bullet}, z_{Xj\bullet} \rangle^2}{d^{2\alpha}} \\
&\leq \frac{\langle c_1, c_2 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, c_1 \rangle^2}{d^{2\alpha}} + \frac{\langle c_2, c_2 \rangle^2}{d^{2\alpha}} + 2\tau_X^2 \sum_{i=3}^d \frac{\|c_1\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2\alpha}} \\
&\quad + 2\tau_X^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2\alpha}} + \tau_X^4 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Xj\bullet}\|_2^2}{d^{2\alpha}}.
\end{aligned}$$

Since $\langle c_1, c_2 \rangle^2$, $\langle c_2, c_1 \rangle^2$ are $O_P(\sqrt{d^\alpha})$, and $\langle c_2, c_2 \rangle^2$ is $O_P(1)$,

$$\frac{\langle c_1, c_2 \rangle^2}{d^{2\alpha}} \xrightarrow[p]{d \rightarrow \infty} 0, \quad \frac{\langle c_2, c_1 \rangle^2}{d^{2\alpha}} \xrightarrow[p]{d \rightarrow \infty} 0, \quad \frac{\langle c_2, c_2 \rangle^2}{d^{2\alpha}} \xrightarrow[p]{d \rightarrow \infty} 0.$$

Note that $\|z_{Xi\bullet}\|_2^2 \sim \chi_n^2$ and that $\|c_1\|_2^2$ and $\|c_2\|_2^2$ are of $O_P(d^\alpha)$. By the law of large numbers,

$$\begin{aligned}
2\tau_X^2 \sum_{i=3}^d \frac{\|c_1\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_X^2}{d^{\alpha-1}} \left\| \frac{d_1}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \\
&\xrightarrow[p]{d \rightarrow \infty} 0 \times O_P(1) \times E(\chi_n^2) = 0, \\
2\tau_X^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_X^2}{d^{\alpha-1}} \left\| \frac{d_2}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \\
&\xrightarrow[p]{d \rightarrow \infty} 0 \times 0 \times E(\chi_n^2) = 0.
\end{aligned}$$

Using the fact that $\alpha > 1$, $\|z_{Xi\bullet}\|_2^2 \sim \chi_n^2$ and $\|z_{Yi\bullet}\|_2^2 \sim \chi_n^2$ and applying the law of large numbers,

$$\begin{aligned}
\tau_X^4 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Xj\bullet}\|_2^2}{d^{2\alpha}} &= \frac{\tau_X^4}{d^{2\alpha-2}} \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \sum_{j=3}^d \frac{\|z_{Xj\bullet}\|_2^2}{d} \\
&\xrightarrow[p]{d \rightarrow \infty} 0 \times n \times n = 0.
\end{aligned}$$

Therefore,

$$\left\| \frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \mathbf{M}_X^{(d)} \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

□

Lemma 8 (CCA HDLSS Asymptotic lemma 8.). *Let \mathbf{M}_X be the matrix defined in Lemma 7. Let $\hat{\lambda}_{X1}^{(d)}$ and $\hat{\xi}_{X1}^{(d)}$ be the first sample eigenvalue and eigenvectors from (27). Then, for $\alpha > 1$,*

$$\left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T - \mathbf{M}_X \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Proof. Using $\mathbf{M}_X^{(d)}$ defined in Lemma 7 and the triangle inequality,

$$\begin{aligned} & \left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T - \mathbf{M}_X^{(d)} \right\|_F \\ &= \left\| \left(\frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \mathbf{M}_X^{(d)} \right) - \left(\frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T \right) \right\|_F \\ &\leq \left\| \frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \mathbf{M}_X^{(d)} \right\|_F + \left\| \frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T \right\|_F. \end{aligned}$$

By Lemma 7,

$$\left\| \frac{n\hat{\Sigma}_X^{(d)}}{d^\alpha} - \mathbf{M}_X^{(d)} \right\|_F^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Write $\mathbf{M}_X^{(d)}$ as,

$$\mathbf{M}_X^{(d)} = \frac{\|c_1\|_2^2}{d^\alpha} e_1^{(d)} \left(e_1^{(d)} \right)^T.$$

Since the first eigenvalue $\hat{\lambda}_{X1}^{(d)}$ and eigenvector $\hat{\xi}_{X1}^{(d)}$ provide the best rank-1 approximation to the matrix $\hat{\lambda}_{X1}^{(d)}/d^\alpha$,

$$\begin{aligned} \left\| \frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} - \frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T \right\|_F^2 &\leq \left\| \frac{\hat{\Sigma}_X^{(d)}}{d^\alpha} - \frac{\|c_1\|_2^2}{nd^\alpha} e_1^{(d)} \left(e_1^{(d)} \right)^T \right\|_F^2 \\ &\leq \left\| \frac{\hat{\Sigma}_X^{(d)}}{d^\alpha} - \frac{\mathbf{M}_X^{(d)}}{n} \right\|_F^2 \\ &\xrightarrow[p]{d \rightarrow \infty} 0. \end{aligned}$$

□

Lemma 9 (CCA HDLSS Asymptotic lemma 9.). *For $\alpha > 1$, the first eigenvalue $\hat{\lambda}_{X1}^{(d)}$ and eigenvectors $\hat{\xi}_{X1}^{(d)}$ converge in probability to the following quantities as $d \rightarrow \infty$,*

$$\frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \xrightarrow[p]{d \rightarrow \infty} C_X, \quad \left\| \hat{\xi}_{X1}^{(d)} - e_1^{(d)} \right\|_2 \xrightarrow[p]{d \rightarrow \infty} 0,$$

where C_X is defined in Lemma 7.

Proof. Let $\tilde{\xi}_{X1}^{(d)}$ be $\hat{\xi}_{X1}^{(d)}$ where their first entries are set to 0. Let $\hat{\xi}_{X1}^{(d)}(i)$ be the i th and j th entries of $\hat{\xi}_{X1}^{(d)}$. By Lemma 8,

$$\begin{aligned} \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \left\| \tilde{\xi}_{X1}^{(d)} \right\|_2^2 \left\| \tilde{\xi}_{X1}^{(d)} \right\|_2^2 &= \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=2}^d \left(\hat{\xi}_{X1}^{(d)}(i) \right)^2 \sum_{j=2}^d \left(\hat{\xi}_{X1}^{(d)}(j) \right)^2 \\ &= \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=2}^d \sum_{j=2}^d \left(\hat{\xi}_{X1}^{(d)}(i) \right)^2 \left(\hat{\xi}_{X1}^{(d)}(j) \right)^2 \\ &= \left\| \frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \tilde{\xi}_{X1}^{(d)} \left(\tilde{\xi}_{X1}^{(d)} \right)^T \right\|_F^2 \\ &\xrightarrow[p]{d \rightarrow \infty} 0. \end{aligned}$$

First, we show that $(\hat{\lambda}_{X1}^{(d)}/d^\alpha)^2 > 0$ in probability. Suppose that $(\hat{\lambda}_{X1}^{(d)}/d^\alpha)^2$ converges in probability to 0. Then, noting that $\left\| \hat{\xi}_{X1}^{(d)} \right\|_2^2 = 1$ and $\left\| \tilde{\xi}_{X1}^{(d)} \right\|_2^2 = 1$,

$$\begin{aligned} \left\| \frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T \right\|_F^2 &= \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=1}^d \sum_{j=1}^d \left(\hat{\xi}_{X1}^{(d)}(i) \right)^2 \left(\hat{\xi}_{X1}^{(d)}(j) \right)^2 \\ &= \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \sum_{i=1}^d \left(\hat{\xi}_{X1}^{(d)} \right)^2 \sum_{j=1}^d \left(\hat{\xi}_{X1}^{(d)}(j) \right)^2 \\ &= \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \left\| \hat{\xi}_{X1}^{(d)} \right\|_2^2 \left\| \hat{\xi}_{X1}^{(d)} \right\|_2^2 \\ &= \left(\frac{\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \right)^2 \\ &\xrightarrow[p]{d \rightarrow \infty} 0, \end{aligned}$$

which contradicts to Lemma 7 (note that the limiting matrix M is not a degenerate matrix).

Therefore,

$$\left\| \tilde{\xi}_{X1}^{(d)} \right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Note that, since the norm of $\hat{\xi}_{X1}^{(d)}$ is 1,

$$\left(\hat{\xi}_{X1}^{(d)}(1) \right)^2 = 1 - \left\| \tilde{\xi}_{X1}^{(d)} \right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 1.$$

Therefore

$$\left\| \hat{\xi}_{X1}^{(d)} - e_1^{(d)} \right\|_2^2 = \left(\hat{\xi}_{X1}^{(d)}(1) - 1 \right)^2 + \left\| \tilde{\xi}_{X1}^{(d)} \right\|_2^2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

To find the limiting value of $\hat{\lambda}_{X1}^{(d)}/d^\alpha$ as $d \rightarrow \infty$, using the unitary invariance property of the Frobenius norm and the previous result about the limiting vectors of $\hat{\xi}_{X1}^{(d)}$ as $d \rightarrow \infty$,

$$\begin{aligned} & \left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T - \mathbf{M}_X^{(d)} \right\|_F^2 \\ &= \left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T - \frac{\|c_1\|_2^2}{d^\alpha} e_1^{(d)} \left(e_1^{(d)} \right)^T \right\|_F^2 \\ &= \left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \left(\hat{\xi}_{X1}^{(d)} \right)^T \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T \hat{\xi}_{X1}^{(d)} - \frac{\|c_1\|_2^2}{d^\alpha} \left(\hat{\xi}_{X1}^{(d)} \right)^T e_1^{(d)} \left(e_1^{(d)} \right)^T \hat{\xi}_{X1}^{(d)} \right\|_F^2 \\ &= \left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} - \frac{\|c_1\|_2^2}{d^\alpha} \left(\hat{\xi}_{X1}^{(d)} \right)^T e_1^{(d)} \left(e_1^{(d)} \right)^T \hat{\xi}_{X1}^{(d)} \right\|_F^2 \\ &= \left(\frac{\|c_1\|_2^2}{d^\alpha} \right)^2 \left\| \frac{n\hat{\lambda}_{X1}^{(d)}/d^\alpha}{\|c_1\|_2^2/d^\alpha} - \left(\hat{\xi}_{X1}^{(d)} \right)^T e_1^{(d)} \left(e_1^{(d)} \right)^T \hat{\xi}_{X1}^{(d)} \right\|_F^2 \\ &\xrightarrow[p]{d \rightarrow \infty} A^2 \left(\frac{B}{C} - 1 \right)^2, \end{aligned}$$

where,

$$A = \lim_{d \rightarrow \infty} \frac{\|c_1\|_2^2}{d^\alpha}, \quad B = \lim_{d \rightarrow \infty} \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha}, \quad C = \lim_{d \rightarrow \infty} \frac{\|c_1\|_2^2}{d^\alpha}.$$

By Lemma 7,

$$\left\| \frac{n\hat{\lambda}_{X1}^{(d)}}{d^\alpha} \hat{\xi}_{X1}^{(d)} \left(\hat{\xi}_{X1}^{(d)} \right)^T - \mathbf{M}_X^{(d)} \right\|_F^2 \xrightarrow{d \rightarrow \infty} 0,$$

and the fact that $A \asymp O_p(1)$,

$$\left(\frac{B}{C} - 1 \right)^2 \xrightarrow{d \rightarrow \infty} 0.$$

Therefore,

$$\frac{B}{C} \xrightarrow{d \rightarrow \infty} 1.$$

□

Lemma 10 (CCA HDLSS Asymptotic lemma 10.). *The sample eigenvalue $\hat{\lambda}_{Xi}^{(d)}$ and eigenvector $\hat{\xi}_{Xi}^{(d)}$, $i = 2, 3, \dots, d$, converge in probability to the following quantities as $d \rightarrow \infty$,*

$$\frac{n\hat{\lambda}_{Xi}^{(d)}}{d} \xrightarrow{d \rightarrow \infty} \tau_X^2, \quad \langle \hat{\xi}_{Xi}^{(d)}, \xi^{(d)} \rangle \xrightarrow{d \rightarrow \infty} 0, \quad i = 2, 3, \dots, n,$$

where $\xi^{(d)}$ is an any given vector in R^d .

Proof. Recall that the underlying random variable $X^{(d)}$ of the sample covariance matrix $\hat{\Sigma}_X^{(d)}$ has a simple spiked covariance structure of (9). Then, the asymptotic behavior of a sample eigenvalue is a direct consequence of Theorem 1 for $\alpha > 1$. The result on the behavior of $\langle \hat{\xi}_{Xi}^{(d)}, \xi^{(d)} \rangle$ is indicated in the proof of Theorem 1 given in [4], even though $\xi^{(d)}$ is fixed at the population counterpart of $\hat{\xi}_{Xi}^{(d)}$ in Theorem 1 for comparison purpose. Note that an inner product of two unit vectors measures a cosine of the angle between the two. Lemma 10 implies that the sample eigenvector $\hat{\xi}_{Xi}^{(d)}$ becomes completely random such that the probability of it being consistent with any given vector is 0 as $d \rightarrow \infty$. □

4.3.3 Behavior of the matrix $\hat{\mathbf{R}}^{(d)}$

The definition of the matrix $\hat{\mathbf{R}}^{(d)}$ is given in (3) with the inverse matrix we are going to use explained in (13). The sample canonical correlation coefficients are found as the singular values of $\hat{\mathbf{R}}^{(d)}$ and sample canonical weight vectors are obtained via unscaling and normalizing the singular vectors of $\hat{\mathbf{R}}^{(d)}$ shown in (4).

$$\begin{aligned}\hat{\mathbf{R}}^{(d)} &= \left(\hat{\boldsymbol{\Sigma}}_X^{(d)}\right)^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}}_{XY}^{(d)} \left(\hat{\boldsymbol{\Sigma}}_Y^{(d)}\right)^{-\frac{1}{2}} \\ &= \sum_{i=1}^n \sqrt{\frac{1}{\hat{\lambda}_{X_i}^{(d)}}} \hat{\xi}_{X_i}^{(d)} \left(\hat{\xi}_{X_i}^{(d)}\right)^T \sum_{i=1}^n \hat{\lambda}_{XY_i}^{(d)} \hat{\eta}_{X_i}^{(d)} \left(\hat{\eta}_{Y_i}^{(d)}\right)^T \sum_{i=1}^n \sqrt{\frac{1}{\hat{\lambda}_{Y_i}^{(d)}}} \hat{\xi}_{Y_i}^{(d)} \left(\hat{\xi}_{Y_i}^{(d)}\right)^T\end{aligned}\quad (28)$$

Consider the two parts of the matrix $\hat{\mathbf{R}}^{(d)}$ in the following,

$$\begin{aligned}& \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\hat{\lambda}_{X2}^{(d)}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\hat{\lambda}_{Xn}^{(d)}}} \end{bmatrix} \begin{bmatrix} \left(\hat{\xi}_{X1}^{(d)}\right)^T \\ \left(\hat{\xi}_{X2}^{(d)}\right)^T \\ \vdots \\ \left(\hat{\xi}_{Xn}^{(d)}\right)^T \end{bmatrix} \begin{bmatrix} \left(\hat{\eta}_{X1}^{(d)}\right)^T \\ \left(\hat{\eta}_{X2}^{(d)}\right)^T \\ \vdots \\ \left(\hat{\eta}_{Xn}^{(d)}\right)^T \end{bmatrix}^T \begin{bmatrix} \sqrt{\hat{\lambda}_{XY1}^{(d)}} & 0 & \dots & 0 \\ 0 & \sqrt{\hat{\lambda}_{XY2}^{(d)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\hat{\lambda}_{XYn}^{(d)}} \end{bmatrix}, \\ & \begin{bmatrix} \sqrt{\hat{\lambda}_{XY1}^{(d)}} & 0 & \dots & 0 \\ 0 & \sqrt{\hat{\lambda}_{XY2}^{(d)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\hat{\lambda}_{XYn}^{(d)}} \end{bmatrix} \begin{bmatrix} \left(\hat{\eta}_{X1}^{(d)}\right)^T \\ \left(\hat{\eta}_{X2}^{(d)}\right)^T \\ \vdots \\ \left(\hat{\eta}_{Xn}^{(d)}\right)^T \end{bmatrix}^T \begin{bmatrix} \left(\hat{\xi}_{Y1}^{(d)}\right)^T \\ \left(\hat{\xi}_{Y2}^{(d)}\right)^T \\ \vdots \\ \left(\hat{\xi}_{Yn}^{(d)}\right)^T \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_{Y1}^{(d)}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\hat{\lambda}_{Y2}^{(d)}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\hat{\lambda}_{Yn}^{(d)}}} \end{bmatrix}.\end{aligned}$$

The first part reduces to the following matrix (call it $\hat{\mathbf{R}}_1^{(d)}$),

$$\hat{\mathbf{R}}_1^{(d)} = \begin{bmatrix} \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle \hat{\xi}_{X1}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle & \frac{\sqrt{\hat{\lambda}_{XY2}^{(d)}}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle \hat{\xi}_{X1}^{(d)}, \hat{\eta}_{X2}^{(d)} \rangle & \dots & \frac{\sqrt{\hat{\lambda}_{XYn}^{(d)}}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle \hat{\xi}_{X1}^{(d)}, \hat{\eta}_{Xn}^{(d)} \rangle \\ \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{X2}^{(d)}}} \langle \hat{\xi}_{X2}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle & \frac{\sqrt{\hat{\lambda}_{XY2}^{(d)}}}{\sqrt{\hat{\lambda}_{X2}^{(d)}}} \langle \hat{\xi}_{X2}^{(d)}, \hat{\eta}_{X2}^{(d)} \rangle & \dots & \frac{\sqrt{\hat{\lambda}_{XYn}^{(d)}}}{\sqrt{\hat{\lambda}_{X2}^{(d)}}} \langle \hat{\xi}_{X1}^{(d)}, \hat{\eta}_{Xn}^{(d)} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{Xn}^{(d)}}} \langle \hat{\xi}_{Xn}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle & \frac{\sqrt{\hat{\lambda}_{XY2}^{(d)}}}{\sqrt{\hat{\lambda}_{Xn}^{(d)}}} \langle \hat{\xi}_{Xn}^{(d)}, \hat{\eta}_{X2}^{(d)} \rangle & \dots & \frac{\sqrt{\hat{\lambda}_{XYn}^{(d)}}}{\sqrt{\hat{\lambda}_{Xn}^{(d)}}} \langle \hat{\xi}_{X1}^{(d)}, \hat{\eta}_{Xn}^{(d)} \rangle \end{bmatrix}. \quad (29)$$

We investigate which value each entry of $\hat{\mathbf{R}}_1^{(d)}(i, j)$ converges in probability to as $d \rightarrow \infty$. Referring to Lemma 3, 6, 9 and 10 on the magnitudes of the sample eigenvalues and singular values, it can be easily noticed that,

$$\begin{aligned} \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} &= \frac{\sqrt{\langle c_1, d_1 \rangle / d^\alpha}}{\sqrt{\langle c_1, c_1 \rangle / d^\alpha}} \asymp O_P(1), \quad \frac{\sqrt{\hat{\lambda}_{XYi}^{(d)}}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} = o_P(1), \quad i, j = 2, 3, \dots, n, \\ \frac{\sqrt{\hat{\lambda}_{XYi}^{(d)}}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} &= O_P(1/\sqrt{d^{\alpha-1}}), \quad \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} = O_P(\sqrt{d^{\alpha-1}}), \quad i = 2, 3, \dots, n. \end{aligned}$$

Then, $\hat{\mathbf{R}}_1^{(d)}$ can be written as,

$$\left\| \left\| \hat{\mathbf{R}}_1^{(d)} - \begin{bmatrix} \frac{\sqrt{\langle c_1, d_1 \rangle}}{\|c_1\|_2} \langle \hat{\xi}_{X1}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle & \mathbf{0}_{1 \times (d-1)} \\ \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{X2}^{(d)}}} \langle \hat{\xi}_{X2}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle & \mathbf{0}_{1 \times (d-1)} \\ \vdots & \vdots \\ \frac{\sqrt{\hat{\lambda}_{XY1}^{(d)}}}{\sqrt{\hat{\lambda}_{Xn}^{(d)}}} \langle \hat{\xi}_{Xn}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle & \mathbf{0}_{1 \times (d-1)} \end{bmatrix} \right\|_2 \right\| \xrightarrow[d \rightarrow \infty]{P} 0.$$

Since $(\sqrt{\hat{\lambda}_{XY1}^{(d)}}/\sqrt{\hat{\lambda}_{Xj}^{(d)}})\langle \hat{\xi}_{Xi}^{(d)}, \hat{\eta}_{X1}^{(d)} \rangle$, for $i = 2, 3, \dots, n$, appears to blow up, we further investigate their magnitudes.

Lemma 11 (CCA HDLSS Asymptotic lemma 11.). *The inner product of the first sample singular vector $\hat{\eta}_{X1}^{(d)}$ and the i th eigenvector $\hat{\xi}_{Xi}^{(d)}$, $i = 2, 3, \dots, n$, converges to 0 with the speed of $o_P(1/\sqrt{d^{\alpha-1}})$ as $d \rightarrow \infty$.*

Proof. Consider $\mathbf{X}^{(d)}$ given in (16), which contains the observations of the random variable $X^{(d)}$.

The singular value decomposition of $\mathbf{X}^{(d)}$ gives,

$$\mathbf{X}_{d \times n}^{(d)} = \begin{bmatrix} \hat{\xi}_{X1}^{(d)} & \hat{\xi}_{X2}^{(d)} & \dots & \hat{\xi}_{Xn}^{(d)} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\hat{\lambda}_{X1}^{(d)}} & 0 & \dots & 0 \\ 0 & \sqrt{\hat{\lambda}_{X2}^{(d)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\hat{\lambda}_{Xn}^{(d)}} \end{bmatrix} \begin{bmatrix} (Z_{\hat{\xi}_{X1}}^{(d)})^T \\ (Z_{\hat{\xi}_{X2}}^{(d)})^T \\ \vdots \\ (Z_{\hat{\xi}_{Xn}}^{(d)})^T \end{bmatrix},$$

where $\hat{\xi}_{Xi}^{(d)}$ and $\hat{\lambda}_{Xi}^{(d)}$ are eigenvectors and eigenvalues of the sample covariance matrix $\hat{\Sigma}_X^{(d)}$, $\|Z_{\hat{\xi}_{Xi}}^{(d)}\| =$

1 and $\langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\xi}_{Xj}}^{(d)} \rangle = 0$ for $i \neq j$. Note that $Z_{\hat{\xi}_{Xi}}^{(d)}$ is the standerzided scores of the projections of the n observations in $\hat{\xi}_{Xi}^{(d)}$ onto the eigenvectors $\hat{\xi}_{Xi}^{(d)}$,

$$Z_{\hat{\xi}_{Xi}}^{(d)} = \frac{(\mathbf{X}^{(d)})^T \hat{\xi}_{Xi}^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}}, \quad i = 1, 2, \dots, n.$$

Define $Z_{\hat{\eta}_{X1}}^{(d)}$ to be the standerzided scores of the projections of the n observations in $\hat{\xi}_{Xi}^{(d)}$ onto the first singular vector $\hat{\eta}_{X1}^{(d)}$ obtained from the sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$,

$$Z_{\hat{\eta}_{X1}}^{(d)} = \frac{(\mathbf{X}^{(d)})^T \hat{\eta}_{X1}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}}.$$

First we show that $Z_{\hat{\xi}_{X1}}^{(d)}$ converges in probability to $Z_{\hat{\eta}_{X1}}^{(d)}$ as $d \rightarrow \infty$. By the triangle inequality,

$$\begin{aligned} \|Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)}\|_2 &= \left\| \frac{(\mathbf{X}^{(d)})^T \hat{\xi}_{X1}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} - \frac{(\mathbf{X}^{(d)})^T \hat{\eta}_{X1}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2 \\ &\leq \left\| \frac{(\mathbf{X}^{(d)})^T (\hat{\xi}_{X1}^{(d)} - e_1^{(d)})}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2 + \left\| \frac{(\mathbf{X}^{(d)})^T (\hat{\eta}_{X1}^{(d)} - e_1^{(d)})}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2. \end{aligned}$$

The first term can be written as,

$$\left\| \frac{(\mathbf{X}^{(d)})^T (\hat{\xi}_{X1}^{(d)} - e_1^{(d)})}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2^2 = \sum_{i=1}^n \left\langle \frac{\mathbf{X}^{(d)}(\bullet, i)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}}, \hat{\xi}_{X1}^{(d)} - e_1^{(d)} \right\rangle^2,$$

where $\mathbf{X}^{(d)}(\bullet, i)$ stands for the i th column of $\mathbf{X}^{(d)}$. Note that the first component in each column of $\mathbf{X}^{(d)}$ is of $O_P(\sqrt{d^\alpha})$, which is that of $\sqrt{\hat{\lambda}_{X1}^{(d)}}$ by Lemma 9, and the rest of the elements are of $O_P(1)$ by Lemma 10. Since $\|\hat{\xi}_{X1}^{(d)} - e_1^{(D)}\|_2$ converge in probability to 0 by Lemma 9, using the Cauchy-Schwarz inequality and the fact that $\alpha > 1$,

$$\begin{aligned} \left\langle \frac{\mathbf{X}^{(d)}(\bullet, i)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}}, \hat{\xi}_{X1}^{(d)} - e_1^{(d)} \right\rangle^2 &\leq \left\| \frac{\mathbf{X}^{(d)}(\bullet, i)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2^2 \|\hat{\xi}_{X1}^{(d)} - e_1^{(d)}\|_2^2 \\ &= \left(\left(\frac{\mathbf{X}^{(d)}(1, i)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^2 + \sum_{j=2}^d \left(\frac{\mathbf{X}^{(d)}(j, i)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^2 \right) \|\hat{\xi}_{X1}^{(d)} - e_1^{(d)}\|_2^2 \end{aligned}$$

$$\begin{aligned}
&= \left(\left(\frac{\mathbf{X}^{(d)}(1, i)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^2 + \frac{d^\alpha}{\hat{\lambda}_{X1}^{(d)}} \frac{1}{d^{\alpha-1}} \sum_{j=2}^d \frac{(\mathbf{X}^{(d)}(j, i))^2}{d} \right) \left\| \hat{\xi}_{X1}^{(d)} - e_1^{(d)} \right\|_2^2 \\
&\xrightarrow{d \rightarrow \infty} 0, \quad i = 1, 2, \dots, n.
\end{aligned}$$

Therefore,

$$\left\| \frac{(\mathbf{X}^{(d)})^T}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} (\hat{\xi}_{X1}^{(d)} - e_1^{(d)}) \right\|_2^2 \xrightarrow{d \rightarrow \infty} 0.$$

Similarly, using the results given in Lemma 3, 9 and 10, we have,

$$\left\| \frac{(\mathbf{X}^{(d)})^T}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} (\hat{\eta}_{X1}^{(d)} - e_1^{(d)}) \right\|_2^2 \xrightarrow{d \rightarrow \infty} 0,$$

which leads to,

$$\left\| Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)} \right\|_2 \xrightarrow{d \rightarrow \infty} 0.$$

Now we show that the inner product $\langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\eta}_{X1}}^{(d)} \rangle$ for $i = 2, 3, \dots, n$, converges to 0 as $d \rightarrow \infty$. Using $\langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\xi}_{X1}}^{(d)} \rangle = 0$ for $i = 2, 3, \dots, n$,

$$\begin{aligned}
\langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\eta}_{X1}}^{(d)} \rangle &= \langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\eta}_{X1}}^{(d)} - Z_{\hat{\xi}_{X1}}^{(d)} + Z_{\hat{\xi}_{X1}}^{(d)} \rangle \\
&= \langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\eta}_{X1}}^{(d)} - Z_{\hat{\xi}_{X1}}^{(d)} \rangle + \langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\xi}_{X1}}^{(d)} \rangle \\
&= \langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\eta}_{X1}}^{(d)} - Z_{\hat{\xi}_{X1}}^{(d)} \rangle.
\end{aligned}$$

Using the Cauchy-Schwarz inequality and the fact that $Z_{\hat{\xi}_{Xi}}^{(d)}$ for $i = 1, 2, \dots, n$, consists of standardized scores (of unit variance),

$$\begin{aligned}
\langle Z_{\hat{\xi}_{Xi}}^{(d)}, Z_{\hat{\eta}_{X1}}^{(d)} - Z_{\hat{\xi}_{X1}}^{(d)} \rangle^2 &\leq \left\| Z_{\hat{\xi}_{Xi}}^{(d)} \right\|_2^2 \left\| Z_{\hat{\eta}_{X1}}^{(d)} - Z_{\hat{\xi}_{X1}}^{(d)} \right\|_2^2 \\
&= 1 \times \left\| Z_{\hat{\eta}_{X1}}^{(d)} - Z_{\hat{\xi}_{X1}}^{(d)} \right\|_2^2 \\
&\xrightarrow{d \rightarrow \infty} 0, \quad i = 1, 2, \dots, n.
\end{aligned}$$

We now show that $\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)} / \sqrt{\hat{\lambda}_{X1}^{(d)}}$ converges in probability to $\hat{\eta}_{X1}^{(d)}$ as $d \rightarrow \infty$. By the triangle inequality,

$$\begin{aligned} \left\| \hat{\eta}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2 &= \left\| \left(\hat{\eta}_{X1}^{(d)} - \hat{\xi}_{X1}^{(d)} \right) - \left(\hat{\xi}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right) \right\|_2 \\ &\leq \left\| \hat{\eta}_{X1}^{(d)} - \hat{\xi}_{X1}^{(d)} \right\|_2 + \left\| \hat{\xi}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2. \end{aligned}$$

As $\|\hat{\xi}_{X1}^{(d)} - e_1^{(d)}\|_2$ and $\|\hat{\eta}_{X1}^{(d)} - e_1^{(d)}\|_2$ converge in probability to 0 by Lemma 3 and 10, so does $\|\hat{\eta}_{X1}^{(d)} - \hat{\xi}_{X1}^{(d)}\|_2$. By the triangle inequality, the second term becomes,

$$\begin{aligned} \left\| \hat{\xi}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2 &= \left\| \left(\hat{\xi}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\xi}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right) - \left(\frac{\mathbf{X}^{(d)} Z_{\hat{\xi}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right) \right\|_2 \\ &\leq \left\| \hat{\xi}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\xi}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2 + \left\| \frac{\mathbf{X}^{(d)} Z_{\hat{\xi}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2. \end{aligned}$$

The first term is 0 since $\mathbf{X}^{(d)} Z_{\hat{\xi}_{X1}}^{(d)} / \sqrt{\hat{\lambda}_{X1}^{(d)}} = \hat{\xi}_{X1}^{(d)}$ in the singular decomposition of $\mathbf{X}^{(d)}$. The second term goes as follows,

$$\begin{aligned} \left\| \frac{\mathbf{X}^{(d)} Z_{\hat{\xi}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2^2 &= \left\| \frac{\mathbf{X}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} (Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)}) \right\|_2^2 \\ &= \sum_{i=1}^d \left\langle \left(\frac{\mathbf{X}^{(d)}(i, \bullet)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^T, Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)} \right\rangle^2, \end{aligned}$$

where $\mathbf{X}^{(d)}(i, \bullet)$ stands for the i th row of the $d \times n$ data matrix $\mathbf{X}^{(d)}$. Note that only the first row of the data matrix $\mathbf{X}^{(d)}$ has its components of magnitude of $O_P(\sqrt{d^\alpha})$, and the components of the rest of the rows are all of $O_P(1)$. Since we showed that $\|Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)}\|_2$ converges in probability to 0 as $d \rightarrow \infty$,

$$\left\langle \left(\frac{\mathbf{X}^{(d)}(1, \bullet)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^T, Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)} \right\rangle^2 \leq \left\| \left(\frac{\mathbf{X}^{(d)}(1, \bullet)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^T \right\|_2^2 \|Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)}\|_2^2$$

$$\xrightarrow[p]{d \rightarrow \infty} 0.$$

Similarly,

$$\begin{aligned} \left\langle \left(\frac{\mathbf{X}^{(d)}(i, \bullet)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^T, Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)} \right\rangle &\leq \left\| \left(\frac{\mathbf{X}^{(d)}(i, \bullet)}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right)^T \right\|_2^2 \left\| Z_{\hat{\xi}_{X1}}^{(d)} - Z_{\hat{\eta}_{X1}}^{(d)} \right\|_2^2 \\ &\xrightarrow[p]{d \rightarrow \infty} 0, \quad i = 2, 3, \dots, d. \end{aligned}$$

Therefore, we proved that,

$$\left\| \hat{\eta}_{X1}^{(d)} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \right\|_2 \xrightarrow[p]{d \rightarrow \infty} 0.$$

Finally to determine the speed of $\langle \hat{\eta}_{X1}^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle$ converging to 0 as $d \rightarrow \infty$,

$$\begin{aligned} \left(Z_{\hat{\eta}_{X1}}^{(d)} \right)^T Z_{\hat{\xi}_{Xi}}^{(d)} &= \frac{\left(Z_{\hat{\eta}_{X1}}^{(d)} \right)^T \left(\mathbf{X}^{(d)} \right)^T \hat{\xi}_{Xi}^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \\ &= \frac{\sqrt{\hat{\lambda}_{X1}^{(d)}} \left(\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)} \right)^T \hat{\xi}_{Xi}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}} \sqrt{\hat{\lambda}_{Xi}^{(d)}}} \\ &= \frac{\sqrt{d^\alpha} \frac{\sqrt{\hat{\lambda}_{X1}^{(d)}}}{\sqrt{d^\alpha}} \frac{\langle \mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle}{\sqrt{\hat{\lambda}_{X1}^{(d)}}}}{\sqrt{d} \frac{\sqrt{\hat{\lambda}_{X1}^{(d)}}}{\sqrt{d}}} , \quad i = 2, 3, \dots, n. \end{aligned}$$

By Lemma 9 and 10, the magnitudes of $\sqrt{\hat{\lambda}_{X1}^{(d)}}/\sqrt{d^\alpha}$ and $\sqrt{\hat{\lambda}_{Xi}^{(d)}}/\sqrt{d}$ are of $O_P(1)$. Recall that $\langle Z_{\hat{\eta}_{X1}}^{(d)}, Z_{\hat{\xi}_{Xi}}^{(d)} \rangle$ converges in probability to 0 as $d \rightarrow \infty$. Hence,

$$\left\langle \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}}, \hat{\xi}_{Xi}^{(d)} \right\rangle \xrightarrow[p]{d \rightarrow \infty} o_P \left(\frac{1}{\sqrt{d^{\alpha-1}}} \right), \quad i = 2, 3, \dots, n.$$

However, we know that,

$$\left\| \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} - \hat{\eta}_{X1}^{(d)} \right\|_2 = \left\| \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} - \frac{\mathbf{X}^{(d)} Z_{\hat{\eta}_{X1}}^{(d)}}{\sqrt{\hat{\lambda}_{XY1}^{(d)}}} \right\|_2 \xrightarrow[d \rightarrow \infty]{p} 0,$$

where the magnitude of the sample singular value $\hat{\lambda}_{XY1}^{(d)}$ is the same as that of $\hat{\lambda}_{X1}^{(d)}$ as shown in Lemma 3. Therefore,

$$\langle \hat{\eta}_{X1}^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle = o_P \left(\frac{1}{\sqrt{d^{\alpha-1}}} \right), \quad i = 2, 3, \dots, n.$$

□

By Lemma 11, the matrix $\hat{\mathbf{R}}_1^{(d)}$ in (29) has the following form as $d \rightarrow \infty$,

$$\left\| \hat{\mathbf{R}}_1^{(d)} - \begin{bmatrix} \frac{\sqrt{\langle c_1, d_1 \rangle}}{\|c_1\|_2} & \mathbf{0}_{1 \times (d-1)} \\ \mathbf{0}_{(d-1) \times 1} & \mathbf{0}_{(d-1) \times (d-1)} \end{bmatrix} \right\|_2 \xrightarrow[d \rightarrow \infty]{P} 0. \quad (30)$$

Similarly, the corresponding part of $\hat{\mathbf{R}}_1^{(d)}$ on the right side of the matrix $\hat{\mathbf{R}}^{(d)}$ (call it $\hat{\mathbf{R}}_2^{(d)}$) reduces to a similar form of (30),

$$\left\| \hat{\mathbf{R}}_2^{(d)} - \begin{bmatrix} \frac{\sqrt{\langle c_1, d_1 \rangle}}{\|c_2\|_2} & \mathbf{0}_{1 \times (d-1)} \\ \mathbf{0}_{(d-1) \times 1} & \mathbf{0}_{(d-1) \times (d-1)} \end{bmatrix} \right\|_2 \xrightarrow[d \rightarrow \infty]{P} 0.$$

Then, the matrix $\hat{\mathbf{R}}^{(d)}$ is written as,

$$\hat{\mathbf{R}}^{(d)} = \begin{bmatrix} \hat{\xi}_{X1}^{(d)} & \hat{\xi}_{X2}^{(d)} & \dots & \hat{\xi}_{Xn}^{(d)} \end{bmatrix} \hat{\mathbf{R}}_1^{(d)} \hat{\mathbf{R}}_2^{(d)} \begin{bmatrix} \left(\hat{\xi}_{Y1}^{(d)} \right)^T \\ \left(\hat{\xi}_{Y2}^{(d)} \right)^T \\ \vdots \\ \left(\hat{\xi}_{Yn}^{(d)} \right)^T \end{bmatrix},$$

and its limiting matrix is,

$$\left\| \hat{\mathbf{R}}^{(d)} - \begin{bmatrix} \hat{\xi}_{X1}^{(d)} & \hat{\xi}_{X2}^{(d)} & \dots & \hat{\xi}_{Xn}^{(d)} \end{bmatrix} \begin{bmatrix} \frac{\langle c_1, d_1 \rangle}{\|c_1\|_2 \|c_2\|_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \left(\hat{\xi}_{Y1}^{(d)} \right)^T \\ \left(\hat{\xi}_{Y2}^{(d)} \right)^T \\ \vdots \\ \left(\hat{\xi}_{Yn}^{(d)} \right)^T \end{bmatrix} \right\|_2 \xrightarrow[d \rightarrow \infty]{P} 0, \quad (31)$$

where $\hat{\xi}_{Yi}^{(d)}$ is a sample eigenvector obtained from the eigendecomposition of the sample covariance matrix shown in (22) from the data matrix $\mathbf{Y}^{(d)}$ of (16).

Perform the SVD of $\hat{\mathbf{R}}^{(d)}$ and denote its sample left and right singular vectors by $\hat{\eta}_{RXi}^{(d)}$ and $\hat{\eta}_{RYi}^{(d)}$ for $i = 1, 2, \dots, n$. From (31), the sample singular vectors of $\hat{\mathbf{R}}^{(d)}$ are linear combinations of $\hat{\xi}_{Xi}^{(d)}$'s or $\hat{\xi}_{Yi}^{(d)}$'s,

$$\hat{\eta}_{RXi}^{(d)} = \sum_{j=1}^n a_j^{(d)} \hat{\xi}_{Xj}^{(d)}, \quad \hat{\eta}_{RYi}^{(d)} = \sum_{k=1}^n b_k^{(d)} \hat{\xi}_{Yk}^{(d)}, \quad i = 1, 2, \dots, n, \quad (32)$$

where $(a_1^{(d)})^2 + (a_2^{(d)})^2 + \dots + (a_n^{(d)})^2 = (b_1^{(d)})^2 + (b_2^{(d)})^2 + \dots + (b_n^{(d)})^2 = 1$ to ensure that norms of $\hat{\eta}_{RXi}^{(d)}$ and $\hat{\eta}_{RYi}^{(d)}$ are of unit length.

4.3.4 Behavior of the first sample canonical weight vector

The sample canonical weight vectors $\hat{\psi}_{Xi}^{(d)}$ and $\hat{\psi}_{Yi}^{(d)}$ are found by unscaling and normalizing $\hat{\eta}_{RX1}^{(d)}$ and $\hat{\eta}_{RY1}^{(d)}$ as in (4) using the pseudoinverse of (13),

$$\hat{\psi}_{Xi}^{(d)} = \frac{\left(\hat{\Sigma}_X^{(d)} \right)^{-\frac{1}{2}} \hat{\eta}_{RXi}^{(d)}}{\left\| \left(\hat{\Sigma}_X^{(d)} \right)^{-\frac{1}{2}} \hat{\eta}_{RXi}^{(d)} \right\|_2}, \quad \hat{\psi}_{Yi}^{(d)} = \frac{\left(\hat{\Sigma}_Y^{(d)} \right)^{-\frac{1}{2}} \hat{\eta}_{RYi}^{(d)}}{\left\| \left(\hat{\Sigma}_Y^{(d)} \right)^{-\frac{1}{2}} \hat{\eta}_{RYi}^{(d)} \right\|_2}, \quad i = 1, 2, \dots, n. \quad (33)$$

To see if $\hat{\psi}_{X1}^{(d)}$ and $\hat{\psi}_{Y1}^{(d)}$ is consistent to their population counterparts $\psi_{X1}^{(d)}$ and $\psi_{Y1}^{(d)}$ given in (7), we investigate the limiting value of their inner products $\langle \hat{\psi}_{X1}^{(d)}, \psi_{X1}^{(d)} \rangle$ and $\langle \hat{\psi}_{Y1}^{(d)}, \psi_{Y1}^{(d)} \rangle$, which measure

the cosine of the angle formed by the two. Expanding (33),

$$\begin{aligned}
\langle \hat{\psi}_{X1}^{(d)}, \psi_{X1}^{(d)} \rangle &= \frac{\left(\psi_{X1}^{(d)} \right)^T \left(\hat{\Sigma}_X^{(d)} \right)^{-\frac{1}{2}} \hat{\eta}_{RX1}^{(d)}}{\left\| \psi_{X1}^{(d)} \right\|_2 \left\| \left(\hat{\Sigma}_X^{(d)} \right)^{-\frac{1}{2}} \hat{\eta}_{RX1}^{(d)} \right\|_2} \\
&= \frac{\left(\psi_{X1}^{(d)} \right)^T \left(\sum_{i=1}^n \hat{\lambda}_{Xi}^{(d)} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T \right)^{-\frac{1}{2}} \hat{\eta}_{RX1}^{(d)}}{\left\| \psi_{X1}^{(d)} \right\|_2 \left\| \left(\sum_{i=1}^n \hat{\lambda}_{Xi}^{(d)} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T \right)^{-\frac{1}{2}} \hat{\eta}_{RX1}^{(d)} \right\|_2} \\
&= \frac{\left(\cos \theta_X e_1^{(d)} + \sin \theta_X e_2^{(d)} \right)^T \left(\sum_{i=1}^n \frac{1}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T \right) \hat{\eta}_{RX1}^{(d)}}{\left\| \cos \theta_X e_1^{(d)} + \sin \theta_X e_2^{(d)} \right\|_2 \left\| \left(\sum_{i=1}^n \frac{1}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T \right) \hat{\eta}_{RX1}^{(d)} \right\|_2}.
\end{aligned}$$

Let's take a close look at the denominator term. It is easy to see that $\left\| \cos \theta_X e_1^{(d)} + \sin \theta_X e_2^{(d)} \right\|_2 = 1$.

Using (32),

$$\begin{aligned}
\left\| \left(\sum_{i=1}^n \frac{1}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T \right) \hat{\eta}_{RX1}^{(d)} \right\|_2 &= \left\| \sum_{i=1}^n \sum_{j=1}^n \frac{a_j}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle \hat{\xi}_{Xi}^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \hat{\xi}_{Xi}^{(d)} \right\|_2 \\
&= \left\| \sum_{i=1}^n \frac{a_i}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \right\|_2.
\end{aligned}$$

Now take a look at the numerator term,

$$\begin{aligned}
&\left(\cos \theta_X e_1^{(d)} + \sin \theta_X e_2^{(d)} \right)^T \left(\sum_{i=1}^n \frac{1}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \left(\hat{\xi}_{Xi}^{(d)} \right)^T \right) \hat{\eta}_{RX1}^{(d)} \\
&= \left(\cos \theta_X e_1^{(d)} + \sin \theta_X e_2^{(d)} \right)^T \left(\sum_{i=1}^n \sum_{j=1}^n \frac{a_j^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle \hat{\xi}_{Xi}^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \hat{\xi}_{Xi}^{(d)} \right) \\
&= \left(\cos \theta_X e_1^{(d)} \right)^T \left(\sum_{i=1}^n \sum_{j=1}^n \frac{a_j^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle \hat{\xi}_{Xi}^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \hat{\xi}_{Xi}^{(d)} \right) + \left(\sin \theta_X e_2^{(d)} \right)^T \left(\sum_{i=1}^n \sum_{j=1}^n \frac{a_j^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle \hat{\xi}_{Xi}^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \hat{\xi}_{Xi}^{(d)} \right) \\
&= \left(\cos \theta_X e_1^{(d)} \right)^T \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \right) + \left(\sin \theta_X e_2^{(d)} \right)^T \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \right)
\end{aligned}$$

$$= \cos \theta_X \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle \right) + \sin \theta_X \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle \right).$$

Hence, we have,

$$\langle \hat{\psi}_{X1}^{(d)}, \psi_{X1}^{(d)} \rangle = \frac{\cos \theta_X \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle \right) + \sin \theta_X \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle \right)}{\left\| \sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} \right\|_2}. \quad (34)$$

Simiarly,

$$\langle \hat{\psi}_{Y1}^{(d)}, \psi_{Y1}^{(d)} \rangle = \frac{\cos \theta_Y \left(\sum_{i=1}^n \frac{b_i^{(d)}}{\sqrt{\hat{\lambda}_{Yi}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Yi}^{(d)} \rangle \right) + \sin \theta_Y \left(\sum_{i=1}^n \frac{b_i^{(d)}}{\sqrt{\hat{\lambda}_{Yi}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{Yi}^{(d)} \rangle \right)}{\left\| \sum_{i=1}^n \frac{b_i^{(d)}}{\sqrt{\hat{\lambda}_{Yi}^{(d)}}} \hat{\xi}_{Yi}^{(d)} \right\|_2}.$$

Observe from (31) that, for $\hat{\eta}_{RX1}^{(d)}$, $a_1^{(d)}$ converges to 1 and $a_i^{(d)}$ for $i = 2, 3, \dots, n$, does to 0 as $d \rightarrow \infty$, which implies,

$$\begin{aligned} & \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle - \frac{1}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{X1}^{(d)} \rangle \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0, \\ & \left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle - \frac{1}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{X1}^{(d)} \rangle \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0, \\ & \left\| \sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} - \frac{1}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \hat{\xi}_{X1}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{P} 0. \end{aligned}$$

Then, (34) becomes,

$$\left(\langle \hat{\psi}_{X1}^{(d)}, \psi_{X1}^{(d)} \rangle - \frac{\cos \theta_X \left(\frac{d^\alpha}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{X1}^{(d)} \rangle \right) + \sin \theta_X \left(\frac{d^\alpha}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{X1}^{(d)} \rangle \right)}{\left\| \frac{d^\alpha}{\sqrt{\hat{\lambda}_{X1}^{(d)}}} \hat{\xi}_{X1}^{(d)} \right\|_2} \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0.$$

Recall from Lemma 9 and 10 that $d^\alpha/\sqrt{\hat{\lambda}_{X1}^{(d)}}$ is of ($\asymp O_P(1)$), $\langle e_1^{(d)}, \hat{\xi}_{X1}^{(d)} \rangle$ and $\langle e_2^{(d)}, \hat{\xi}_{X1}^{(d)} \rangle$ converge in probability to 1 and 0, respectively. Therefore, we finally have,

$$\left(\langle \hat{\psi}_{X1}^{(d)}, \psi_{X1}^{(d)} \rangle - \cos \theta_X \right)^2 \xrightarrow{d \rightarrow \infty} 0.$$

Similarly for $\hat{\psi}_{Y1}^{(d)}$,

$$\left(\langle \hat{\psi}_{Y1}^{(d)}, \psi_{Y1}^{(d)} \rangle - \cos \theta_Y \right)^2 \xrightarrow{d \rightarrow \infty} 0.$$

4.3.5 Behavior of the first sample canonical correlation coefficient

The limiting value of first sample canonical correlation coefficients is found as the first singular value of the matrix $\hat{\mathbf{R}}^{(d)}$ in (31) under the limiting operation of $d \rightarrow \infty$, which turns out its (1,1)th entry,

$$\left(\hat{\rho}_1 - \frac{\langle c_1, d_1 \rangle / d^\alpha}{\|c_1\|_2^2 / d^\alpha \|c_2\|_2^2 / d^\alpha} \right)^2 \xrightarrow{d \rightarrow \infty} 0.$$

Recall from Lemma 1 and 7 that $\langle c_1, d_1 \rangle / d^\alpha$ is the limiting value of (1,1) entry of $n\hat{\Sigma}_{XY}^{(d)} / d^\alpha$ as $d \rightarrow \infty$ and that $\|c_1\|_2^2 / d^\alpha$ and $\|c_2\|_2^2 / d^\alpha$ are the limiting values of the (1,1) entries of $n\hat{\Sigma}_X^{(d)} / d^\alpha$ and $n\hat{\Sigma}_Y^{(d)} / d^\alpha$. That is, those values are found in the (d+1,d+1), (1,1) and (1,d+1) positions of the matrix of (21) scaled by d^α as d increases,

$$\frac{1}{d^\alpha} \begin{bmatrix} n\hat{\Sigma}_X^{(d)} & n\hat{\Sigma}_{XY}^{(d)} \\ \left(n\hat{\Sigma}_{XY}^{(d)}\right)^T & n\hat{\Sigma}_Y^{(d)} \end{bmatrix} = \frac{1}{d^\alpha} \begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix} \begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix}^T.$$

Referring to (16) and (17), we switch the (1,2) and (1,d+1) entries, the (2,1) and (d+1,1) entries and the (2,2) and (d+1,d+1) entries of the matrix $\Sigma_T^{(2d)}$ and denote the resulting matrix by $\acute{\Sigma}_T^{(2d)}$. We, also, switch the second and $(d+1)$ th rows of the matrix $\mathbf{Z}^{(2d)}$ and denote the resulting matrix by $\acute{\mathbf{Z}}^{(2d)}$. Then, the two data matrices below,

$$\begin{bmatrix} \mathbf{X}^{(d)} \\ \mathbf{Y}^{(d)} \end{bmatrix} = \left(\Sigma_T^{(2d)} \right)^{\frac{1}{2}} \mathbf{Z}^{(2d)} \text{ and } \left(\acute{\Sigma}_T^{(2d)} \right)^{\frac{1}{2}} \acute{\mathbf{Z}}^{(2d)},$$

are the same only except that the $(d+1)$ th row of the first matrix is the second row of the second one. Following the similar steps as done in the proof of Lemma 1, it can be shown that,

$$\left\| \left(\frac{\dot{\Sigma}_T^{(2d)}}{d^\alpha} \right)^{\frac{1}{2}} \dot{\mathbf{Z}}^{(2d)} - \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y & \mathbf{0}_{1 \times (2d-2)} \\ \rho\sigma_X\sigma_Y & \sigma_X^2 & \mathbf{0}_{1 \times (2d-2)} \\ \mathbf{0}_{(2d-2) \times 1} & \mathbf{0}_{(2d-2) \times 1} & \mathbf{0}_{(2d-2) \times (2d-2)} \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} z_{X1\bullet} \\ z_{Y1\bullet} \\ z_{X2\bullet} \\ \vdots \\ z_{Xd\bullet} \\ z_{Y2\bullet} \\ \vdots \\ z_{Yd\bullet} \end{bmatrix} \right\|_2 \xrightarrow[d \rightarrow \infty]{P} 0.$$

Then, the limiting values of $\langle c_1, d_1 \rangle / d^\alpha$, $\|c_1\|_2^2 / d^\alpha$ and $\|c_1\|_2^2 / d^\alpha$ are found in the first 2×2 block of the following matrix, as $d \rightarrow \infty$,

$$\frac{1}{d^\alpha} \left(\dot{\Sigma}_T^{(2d)} \right)^{\frac{1}{2}} \dot{\mathbf{Z}}^{(2d)} \left(\dot{\mathbf{Z}}^{(2d)} \right)^T \left(\dot{\Sigma}_T^{(2d)} \right)^{\frac{1}{2}} = \left(\frac{\dot{\Sigma}_T^{(2d)}}{d^\alpha} \right)^{\frac{1}{2}} \dot{\mathbf{Z}}^{(2d)} \left(\dot{\mathbf{Z}}^{(2d)} \right)^T \left(\frac{\dot{\Sigma}_T^{(2d)}}{d^\alpha} \right)^{\frac{1}{2}}.$$

Therefore, the limiting values of $\langle c_1, d_1 \rangle / d^\alpha$, $\|c_1\|_2^2 / d^\alpha$ and $\|c_1\|_2^2 / d^\alpha$ are identified as the (2,2), (1,1) and (2,1) entries of the following matrix, respectively,

$$\begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} z_{X1\bullet} \\ z_{Y1\bullet} \end{bmatrix} \begin{bmatrix} z_{X1\bullet} \\ z_{Y1\bullet} \end{bmatrix}^T \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix}^{\frac{1}{2}}.$$

4.3.6 Behavior of the rest of sample canonical weight vectors

We use the equation (34) for the second canonical weight vector $\hat{\psi}_{X2}^{(d)}$. Recall from (31) that $a_1^{(d)}$ of $\hat{\eta}_{RXi}^{(d)}$, for $i = 2, 3, \dots, n$, converges to 0, which implies,

$$\left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle - \sum_{j=2}^n \frac{a_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0,$$

$$\left(\sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle - \sum_{j=2}^n \frac{a_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \langle e_2^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0,$$

$$\left\| \sum_{i=1}^n \frac{a_i^{(d)}}{\sqrt{\hat{\lambda}_{Xi}^{(d)}}} \hat{\xi}_{Xi}^{(d)} - \sum_{j=2}^n \frac{a_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \hat{\xi}_{Xj}^{(d)} \right\|_2^2 \xrightarrow{d \rightarrow \infty} 0.$$

Then, (34) becomes,

$$\left(\langle \hat{\psi}_{X2}^{(d)}, \psi_{X1}^{(d)} \rangle - \frac{\cos \theta_X \left(\sum_{j=2}^n \frac{da_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \right) + \sin \theta_X \left(\sum_{j=2}^n \frac{da_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \right)}{\left\| \sum_{j=2}^n \frac{da_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \hat{\xi}_{Xj}^{(d)} \right\|_2} \right)^2 \xrightarrow{d \rightarrow \infty} 0.$$

We know from 10 that $d/\sqrt{\hat{\lambda}_{Xi}^{(d)}}$, for $i = 2, 3, \dots, n$, is of ($\asymp O_P(1)$) and that $\langle e_1^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle$ and $\langle e_2^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle$, for $i = 2, 3, \dots, n$, both converge in probability to 0. Therefore, we finally have,

$$\left(\langle \hat{\psi}_{Xi}^{(d)}, \psi_{X1}^{(d)} \rangle - 0 \right)^2 \xrightarrow{d \rightarrow \infty} 0, \quad i = 2, 3, \dots, n.$$

Results are similar for $\hat{\psi}_{Xi}^{(d)}$ for $i = 3, 4, \dots, n$. By the similar argument,

$$\left(\langle \hat{\psi}_{Yi}^{(d)}, \psi_{Y1}^{(d)} \rangle - 0 \right)^2 \xrightarrow{d \rightarrow \infty} 0, \quad i = 2, 3, \dots, n.$$

4.3.7 Behavior of the rest of sample canonical correlation coefficients

The asymptotic i th sample canonical correlation coefficients, for $i = 2, 3, \dots, n$, is found as the i th singular value of the matrix $\hat{\mathbf{R}}^{(d)}$ in (31) under the limiting operation of $d \rightarrow \infty$,

$$\hat{\rho}_i \xrightarrow{d \rightarrow \infty} 0, \quad i = 1, 2, \dots, n.$$

4.4 Case of $\alpha < 1$

We, now, prove Theorem 1 under the condition of $\alpha < 1$ with the spiked model of the population covariance structure of $X^{(d)}$ and $Y^{(d)}$ described in (9).

4.4.1 Behavior of the cross-covariance matrix

Lemma 12 (CCA HDLSS Asymptotic lemma 12.). *For $\alpha < 1$, the magnitudes of the sample singular values $\hat{\lambda}_{XYi}^{(d)}$ for $i = 1, 2, \dots, n$ are of $O_P(d)$ as $d \rightarrow \infty$.*

Proof. Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\left\| \frac{n\hat{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 &= \frac{\langle c_1, d_1 \rangle^2}{d^{2-2\alpha}d^{2\alpha}} + \frac{\langle c_1, d_2 \rangle^2}{d^{2-\alpha}d^\alpha} + \frac{\langle c_2, d_1 \rangle^2}{d^{2-\alpha}d^\alpha} + \frac{\langle c_2, d_2 \rangle^2}{d^2} \\
&\quad + \tau_Y^2 \sum_{i=3}^d \frac{\langle c_1, z_{Yi\bullet} \rangle^2}{d^{2-\alpha}d^\alpha} + \tau_Y^2 \sum_{i=3}^d \frac{\langle c_2, z_{Yi\bullet} \rangle^2}{d^2} + \tau_X^2 \sum_{i=3}^d \frac{\langle z_{Xi\bullet}, d_1 \rangle^2}{d^{2-\alpha}d^\alpha} \\
&\quad + \tau_X^2 \sum_{i=3}^d \frac{\langle z_{Xi\bullet}, d_2 \rangle^2}{d^2} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\langle z_{Xi\bullet}, z_{Yj\bullet} \rangle^2}{d^2} \\
&\leq \frac{\langle c_1, d_1 \rangle^2}{d^{2-2\alpha}d^{2\alpha}} + \frac{\langle c_1, d_2 \rangle^2}{d^{2-\alpha}d^\alpha} + \frac{\langle c_2, d_1 \rangle^2}{d^{2-\alpha}d^\alpha} + \frac{\langle c_2, d_2 \rangle^2}{d^2} \\
&\quad + \tau_Y^2 \sum_{i=3}^d \frac{\|c_1\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^{2-\alpha}d^\alpha} + \tau_Y^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^2} + \tau_X^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|d_1\|_2^2}{d^{2-\alpha}d^\alpha} \\
&\quad + \tau_X^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|d_2\|_2^2}{d^2} + \tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Yj\bullet}\|_2^2}{d^2}.
\end{aligned} \tag{35}$$

Since $\langle c_1, d_1 \rangle^2$ is of $O_P(d^\alpha)$, $\langle c_1, d_2 \rangle^2$ and $\langle c_2, d_1 \rangle^2$ are of $O_P(\sqrt{d^\alpha})$, and $\langle c_2, d_2 \rangle^2$ is of $O_P(1)$, using $\alpha < 1$,

$$\frac{\langle c_1, d_1 \rangle^2}{d^{2-2\alpha}d^{2\alpha}} \xrightarrow[p]{d \rightarrow \infty} 0, \quad \frac{\langle c_1, d_2 \rangle^2}{d^{2-\alpha}d^\alpha} \xrightarrow[p]{d \rightarrow \infty} 0, \quad \frac{\langle c_2, d_1 \rangle^2}{d^{2-\alpha}d^\alpha} \xrightarrow[p]{d \rightarrow \infty} 0, \quad \frac{\langle c_2, d_2 \rangle^2}{d^2} \xrightarrow[p]{d \rightarrow \infty} 0.$$

It is not hard to see that $\|z_{Xi\bullet}\|_2^2 \sim \chi_n^2$ and $\|z_{Yi\bullet}\|_2^2 \sim \chi_n^2$. Note that $\|c_1\|_2^2$ and $\|d_1\|_2^2$ are of $O_P(d^\alpha)$ and that $\|c_2\|_2^2$ and $\|d_2\|_2^2$ are of $O_P(1)$. With the law of large numbers and the condition that $\alpha < 1$,

$$\begin{aligned}
\tau_Y^2 \sum_{i=3}^d \frac{\|c_1\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^{2-\alpha}d^\alpha} &= \frac{\tau_Y^2}{d^{1-\alpha}} \left\| \frac{c_1}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Yi\bullet}\|_2^2}{d} \\
&\xrightarrow[p]{d \rightarrow \infty} 0 \times O_P(1) \times E(\chi_n^2) = 0, \\
\tau_Y^2 \sum_{i=3}^d \frac{\|c_2\|_2^2 \|z_{Yi\bullet}\|_2^2}{d^2} &= \frac{\tau_Y^2 \|c_2\|_2^2}{d} \sum_{i=3}^d \frac{\|z_{Yi\bullet}\|_2^2}{d} \\
&\xrightarrow[p]{d \rightarrow \infty} 0 \times E(\chi_n^2) = 0, \\
\tau_X^2 \sum_{i=3}^d \frac{\|d_1\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^{2-\alpha}d^\alpha} &= \frac{\tau_X^2}{d^{1-\alpha}} \left\| \frac{d_1}{\sqrt{d^\alpha}} \right\|_2^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \\
&\xrightarrow[p]{d \rightarrow \infty} 0 \times O_P(1) \times E(\chi_n^2) = 0,
\end{aligned}$$

$$\begin{aligned}\tau_X^2 \sum_{i=3}^d \frac{\|d_2\|_2^2 \|z_{Xi\bullet}\|_2^2}{d^2} &= \frac{\tau_X^2 \|d_2\|_2^2}{d} \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \\ &\xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} 0 \times E(\chi_n^2) = 0.\end{aligned}$$

Using $\alpha > 1$ and the law of large numbers,

$$\begin{aligned}\tau_X^2 \tau_Y^2 \sum_{i=3}^d \sum_{j=3}^d \frac{\|z_{Xi\bullet}\|_2^2 \|z_{Yj\bullet}\|_2^2}{d^2} &= \tau_X^2 \tau_Y^2 \sum_{i=3}^d \frac{\|z_{Xi\bullet}\|_2^2}{d} \sum_{j=3}^d \frac{\|z_{Yj\bullet}\|_2^2}{d} \\ &\xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} \tau_X^2 \tau_Y^2 \times E(\chi_n^2) \times E(\chi_n^2) \\ &= \tau_X^2 \tau_Y^2 n^2.\end{aligned}$$

We have,

$$\left\| \frac{\hat{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 = \sum_{i=1}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} \leq \tau_X^2 \tau_Y^2.$$

Therefore, the magnitudes of the sample singular values $\hat{\lambda}_{XYi}^{(d)}$ for $i = 1, 2, \dots, n$ are of $O_p(d)$. \square

Lemma 13 (CCA HDLSS Asymptotic lemma 13.). *This lemma improves Lemma 12. For $\alpha < 1$, the sample singular values $n\hat{\lambda}_{XYi}^{(d)}$ for $i = 1, 2, \dots, n$ are of $(\asymp O_p(d))$ with the following limiting quantity as $d \rightarrow \infty$,*

$$\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \xrightarrow[p \rightarrow \infty]{d \rightarrow \infty} = \frac{\tau_X \tau_Y}{n}, \quad i = 1, 2, \dots, n.$$

Proof. First we calculate the exact limiting value of the Frobenius norm of $\hat{\Sigma}_{XY}^{(d)}/d$ as $d \rightarrow \infty$.

In (35), the first 8 terms converges in probability to 0 as $d \rightarrow \infty$. The last term of (35) expands as,

$$\begin{aligned}\sum_{i=3}^d \sum_{j=3}^d \frac{\langle z_{Xi\bullet}, z_{Yj\bullet} \rangle^2}{d^2} &= \sum_{i=3}^d \sum_{j=3}^d \left(\sum_{k=1}^n \frac{z_{Xik} z_{Yjk}}{d} \right)^2 \\ &= \sum_{k=1}^n \left(\sum_{i=3}^d \frac{z_{Xik}^2}{d} \right) \left(\sum_{j=3}^d \frac{z_{Yjk}^2}{d} \right) \\ &\quad + \sum_{k \neq k'} \left(\sum_{i=3}^d \frac{z_{Xik} z_{Yik'}}{d} \right) \left(\sum_{j=3}^d \frac{z_{Xjk} z_{Yjk'}}{d} \right).\end{aligned}$$

Since $E(z_{yik}z_{xik'}) = 0$ for $k \neq k'$, by the law of large numbers,

$$\sum_{k \neq k'} \left(\sum_{i=3}^d \frac{z_{yik}z_{xik'}}{d} \right) \left(\sum_{j=3}^d \frac{z_{yjk}z_{xjk'}}{d} \right) \xrightarrow[d \rightarrow \infty]{P} 0.$$

Since z_{yik}^2 and z_{xik}^2 follow a Chi-squared distribution with a degree of freedom of 1, by the law of large numbers,

$$\sum_{k=1}^n \left(\sum_{i=3}^d \frac{z_{yik}^2}{d} \right) \left(\sum_{i=3}^d \frac{z_{xik}^2}{d} \right) \xrightarrow[d \rightarrow \infty]{P} n \times 1 \times 1 = n.$$

Hence,

$$\left\| \frac{n \hat{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 \xrightarrow[d \rightarrow \infty]{P} n \tau_X^2 \tau_Y^2. \quad (36)$$

Now we set a bound for each sample singular value $\hat{\lambda}_{XYi}^{(d)}/d$ for $i = 1, 2, \dots, n$ as $d \rightarrow \infty$. Consider the data matrix $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$ in (19). Write the singular vectors $\hat{\eta}_{Xi}^{(d)}$ and $\hat{\eta}_{Yi}^{(d)}$ of the sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$ of (22) by the linear combinations of the sample eigenvectors $\hat{\xi}_{Xj}^{(d)}$ and $\hat{\xi}_{Yj}^{(d)}$,

$$\hat{\eta}_{Xi}^{(d)} = \sum_{j=1}^n a_j \hat{\xi}_{Xj}^{(d)}, \quad \hat{\eta}_{Yi}^{(d)} = \sum_{j=1}^n b_j \hat{\xi}_{Yj}^{(d)}, \quad i = 1, 2, \dots, n,$$

where $a_1^2 + a_2^2 + \dots + a_n^2 = b_1^2 + b_2^2 + \dots + b_n^2 = 1$ to ensure a unit length. Using $Cov^2(X, Y) \leq Var(X)Var(Y)$,

$$\begin{aligned} \left(\hat{\lambda}_{XYi}^{(d)} \right)^2 &= \frac{1}{n^2} \left\langle \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)}, \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)} \right\rangle^2 \\ &\leq \frac{1}{n^2} \left\langle \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)} \right\rangle \left\langle \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)}, \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)} \right\rangle. \end{aligned}$$

Using the fact that the covariance between two sets of scores of the projections of the observations onto any two different eigenvectors is 0,

$$\frac{1}{n} \left\langle \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)} \right\rangle = \frac{1}{n} \left\langle \left(\mathbf{X}^{(d)} \right)^T \sum_{j=1}^n a_j \hat{\xi}_{Xj}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T \sum_{j=1}^n a_j \hat{\xi}_{Xj}^{(d)} \right\rangle$$

$$\begin{aligned}
&= \sum_{j=j'} \frac{1}{n} \left\langle \left(\mathbf{X}^{(d)} \right)^T a_j \hat{\xi}_{Xj}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T a_{j'} \hat{\xi}_{Xj'}^{(d)} \right\rangle \\
&\quad + \sum_{j \neq j'} \frac{1}{n} \left\langle \left(\mathbf{X}^{(d)} \right)^T a_j \hat{\xi}_{Xj}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T a_{j'} \hat{\xi}_{Xj'}^{(d)} \right\rangle \\
&= \sum_{j=j'} \frac{1}{n} \left\langle \left(\mathbf{X}^{(d)} \right)^T a_j \hat{\xi}_{Xj}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T a_{j'} \hat{\xi}_{Xj'}^{(d)} \right\rangle \\
&= \sum_{j=1}^n \frac{a_j^2}{n} \left\langle \left(\mathbf{X}^{(d)} \right)^T \hat{\xi}_{Xj}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T \hat{\xi}_{Xj}^{(d)} \right\rangle \\
&= \sum_{j=1}^n a_j^2 \hat{\lambda}_{Xj}^{(d)}.
\end{aligned}$$

Similarly,

$$\frac{1}{n} \left\langle \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)}, \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)} \right\rangle = \sum_{k=1}^n b_k^2 \hat{\lambda}_{Yk}^{(d)}.$$

By Lemma 10,

$$\begin{aligned}
\left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 &\leq \frac{1}{n^2} \frac{\left\langle \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)}, \left(\mathbf{X}^{(d)} \right)^T \hat{\eta}_{Xi}^{(d)} \right\rangle}{d} \frac{\left\langle \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)}, \left(\mathbf{Y}^{(d)} \right)^T \hat{\eta}_{Yi}^{(d)} \right\rangle}{d} \\
&= \left(\sum_{j=1}^n a_j^2 \frac{\hat{\lambda}_{Xj}^{(d)}}{d} \right) \left(\sum_{k=1}^n b_k^2 \frac{\hat{\lambda}_{Yk}^{(d)}}{d} \right) \\
&\xrightarrow[d \rightarrow \infty]{P} \frac{\tau_X^2}{n} \frac{\tau_Y^2}{n} \left(\sum_{j=1}^n a_j^2 \right) \left(\sum_{k=1}^n b_k^2 \right) \\
&= \frac{\tau_X^2 \tau_Y^2}{n^2}, \quad i = 1, 2, \dots, n.
\end{aligned}$$

The constraints above and the proven fact of (36),

$$\left\| \frac{\hat{\Sigma}_{XY}^{(d)}}{d} \right\|_F^2 = \sum_{i=1}^n \left(\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \right)^2 \xrightarrow[d \rightarrow \infty]{P} \frac{1}{n} \tau_X^2 \tau_Y^2,$$

implies that,

$$\frac{\hat{\lambda}_{XYi}^{(d)}}{d} \xrightarrow[d \rightarrow \infty]{P} \frac{\tau_X \tau_Y}{n}, \quad i = 1, 2, \dots, n.$$

□

4.4.2 Behavior of the sample covariance matrices

Here, we only include the result about the sample covariance matrix $\hat{\Sigma}_X^{(d)}$. That of $\hat{\Sigma}_Y^{(d)}$ is similar.

Lemma 14 (CCA HDLSS Asymptotic lemma 14.). *The sample eigenvalue $\hat{\lambda}_{Xi}^{(d)}$ and eigenvector $\hat{\xi}_{Xi}^{(d)}$, $i = 2, 3, \dots, d$, converge in probability to the following quantities as $d \rightarrow \infty$,*

$$\frac{\hat{\lambda}_{Xi}^{(d)}}{d} \xrightarrow[d \rightarrow \infty]{P} \frac{\tau_X^2}{n}, \quad \langle \hat{\xi}_{Xi}^{(d)}, \xi^{(d)} \rangle \xrightarrow[d \rightarrow \infty]{P} 0, \quad i = 1, 2, \dots, n,$$

where $\xi^{(d)}$ is an any given vector in R^d .

Proof. Recall that the underlying random variable $X^{(d)}$ of the sample covariance matrix $\hat{\Sigma}_X^{(d)}$ has a simple spiked covariance structure of (9). Then, the asymptotic behavior of a sample eigenvalue is a direct consequence of Theorem 1 for $\alpha < 1$. The result on the behavior of $\langle \hat{\xi}_{Xi}^{(d)}, \xi^{(d)} \rangle$ is indicated in the proof of Theorem 1 given in [4]. □

4.4.3 Behavior of the matrix $\hat{\mathbf{R}}^{(d)}$

Lemma 13 states that the the singular values of the sample cross-covariance matrix $\hat{\Sigma}_{XY}^{(d)}$ become indistinguishable in the limit of $d \rightarrow \infty$, in which situation there exist an infinite number of choices for a singular vector set. We choose its singular vectors such that the sample eigenvectors of the sample covariance matrices $\hat{\Sigma}_X^{(d)}$ and $\hat{\Sigma}_Y^{(d)}$ are their limiting quantities,

$$\left\| \hat{\eta}_{Xi}^{(d)} - \hat{\xi}_{Xi}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{P} 0, \quad \left\| \hat{\eta}_{Yj}^{(d)} - \hat{\xi}_{Yj}^{(d)} \right\|_2^2 \xrightarrow[d \rightarrow \infty]{P} 0, \quad i, j = 1, 2, \dots, n.$$

Then, it is easily shown that the matrix $\hat{\mathbf{R}}^{(d)}$ defined in (28), reduces to the following limiting form, using the limiting values of the sample singular vectors and eigenvectors given in Lemma 13 and 14,

$$\left\| \hat{\mathbf{R}}^{(d)} - \begin{bmatrix} \hat{\xi}_{X1}^{(d)} & \hat{\xi}_{X2}^{(d)} & \dots & \hat{\xi}_{Xn}^{(d)} \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \left(\hat{\xi}_{Y1}^{(d)} \right)^T \\ \left(\hat{\xi}_{Y2}^{(d)} \right)^T \\ \vdots \\ \left(\hat{\xi}_{Yn}^{(d)} \right)^T \end{bmatrix} \right\|_F^2 \xrightarrow[d \rightarrow \infty]{P} 0. \quad (37)$$

4.4.4 Behavior of sample canonical weight vectors

Use the equation (34) for a sample canonical weight vector $\hat{\psi}_{Xi}^{(d)}$ represented as in (32) for a given i . We know from Lemma 14 that the magnitude of $d/\sqrt{\hat{\lambda}_{Xi}^{(d)}}$, for $i = 1, 2, \dots, n$, is of ($\asymp O_P(1)$) and that $\langle e_1^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle$ and $\langle e_2^{(d)}, \hat{\xi}_{Xi}^{(d)} \rangle$, for $i = 1, 2, \dots, n$, converge in probability to 0 as $d \rightarrow \infty$, which leads to,

$$\left(\langle \hat{\psi}_{Xi}^{(d)}, \psi_{X1}^{(d)} \rangle - \frac{\cos \theta_X \left(\sum_{j=1}^n \frac{da_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \right) + \sin \theta_X \left(\sum_{j=1}^n \frac{da_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \langle e_1^{(d)}, \hat{\xi}_{Xj}^{(d)} \rangle \right)}{\left\| \sum_{j=1}^n \frac{da_j^{(d)}}{\sqrt{\hat{\lambda}_{Xj}^{(d)}}} \hat{\xi}_{Xj}^{(d)} \right\|_2} \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0.$$

Hence, we have,

$$\left(\langle \hat{\psi}_{Xi}^{(d)}, \psi_{X1}^{(d)} \rangle - 0 \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0, \quad i = 1, 2, \dots, n.$$

Similarly,

$$\left(\langle \hat{\psi}_{Yi}^{(d)}, \psi_{Y1}^{(d)} \rangle - 0 \right)^2 \xrightarrow[d \rightarrow \infty]{P} 0, \quad i = 1, 2, \dots, n.$$

4.4.5 Behavior of sample canonical correlation coefficients

The asymptotic i th sample canonical correlation coefficients is found as the i th singular value of the matrix (37) under the limiting operation of $d \rightarrow \infty$,

$$\hat{\rho}_i \xrightarrow[d \rightarrow \infty]{P} 1, \quad i = 1, 2, \dots, n.$$

5 Simulation

Simulation study in this section aims at verifying the asymptotic behavior of sample canonical correlation coefficients and their corresponding weight vectors given in the main theorem 1 as dimension d grows with sample size n fixed. We first state the parameter settings to be used. For the spiked covariance structures of the random variables $X^{(d)}$ and $Y^{(d)}$ described in (5) and (6), we set $\sigma_X^2 = \tau_X^{(d)} = \sigma_Y^2 = \tau_Y^{(d)} = 1$. The population canonical weight vectors described in (7) and

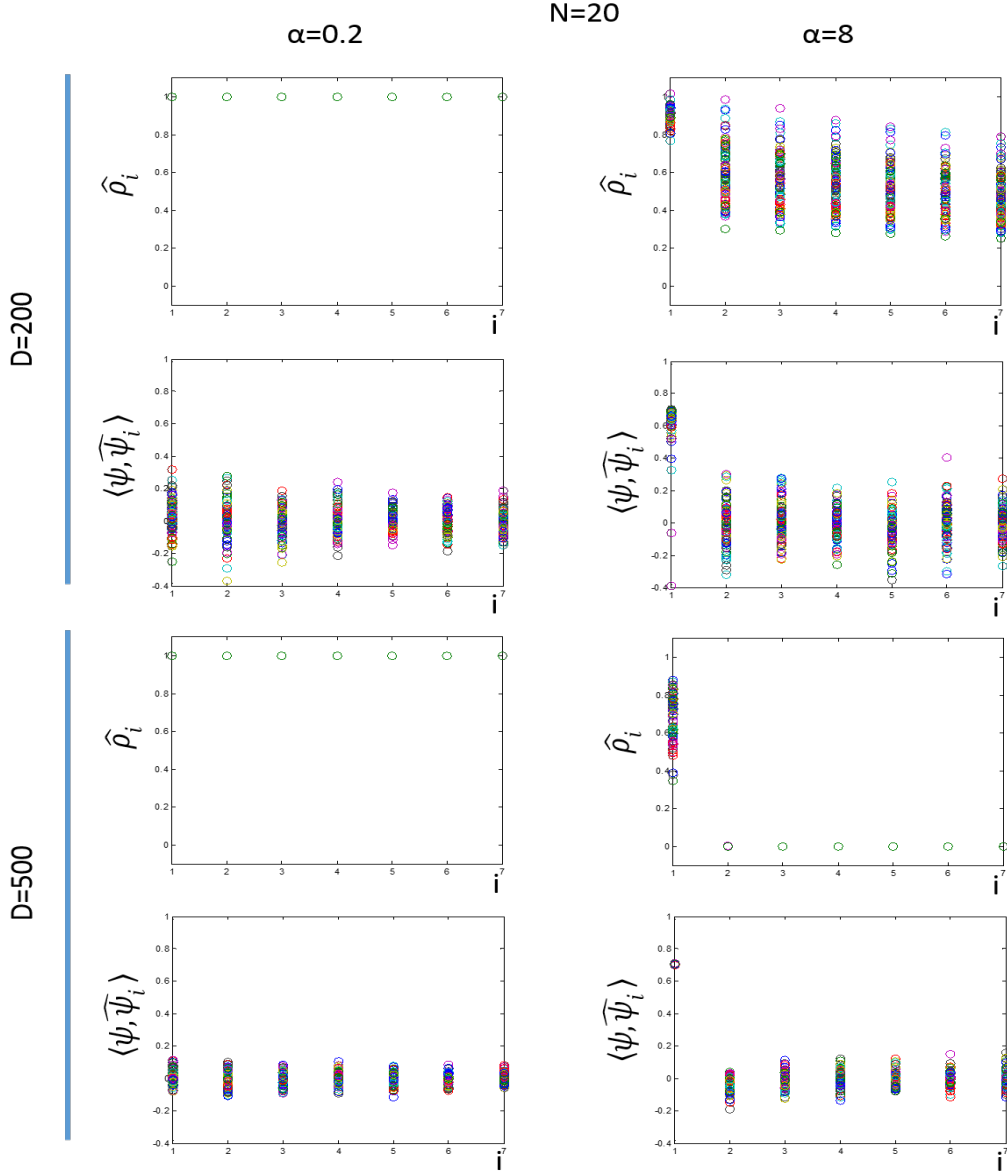


Figure 1. Estimated sample canonical correlation coefficients $\hat{\rho}_i^{(d)}$ and inner products of the sample left canonical weight vectors $\hat{\psi}_{X_i}^{(d)}$ and the population canonical weight vector $\psi_{X_i}^{(d)}$, for $i = 1, 2, \dots, 5$, obtained from 100 repetitions of simulations for different settings of dimension d and exponent α with a sample size of $n = 20$.

population canonical correlation coefficient are set to be,

$$\psi_X^{(d)} = (\cos 0.75\pi)e_1^{(d)} + (\sin 0.75\pi)e_2^{(d)}, \quad \psi_Y^{(d)} = (\cos 0.75\pi)e_1^{(d)} + (\sin 0.75\pi)e_2^{(d)}, \quad \rho = 0.7.$$

Note that $\langle \psi_X^{(d)}, e_1^{(d)} \rangle = \langle \psi_Y^{(d)}, e_1^{(d)} \rangle = \cos 0.75\pi = 0.7071$, which implies that the angle between $\psi_X^{(d)}$ and $e_1^{(d)}$ is 135° . The population cross-covariance structure of $X^{(d)}$ and $Y^{(d)}$ can be accordingly

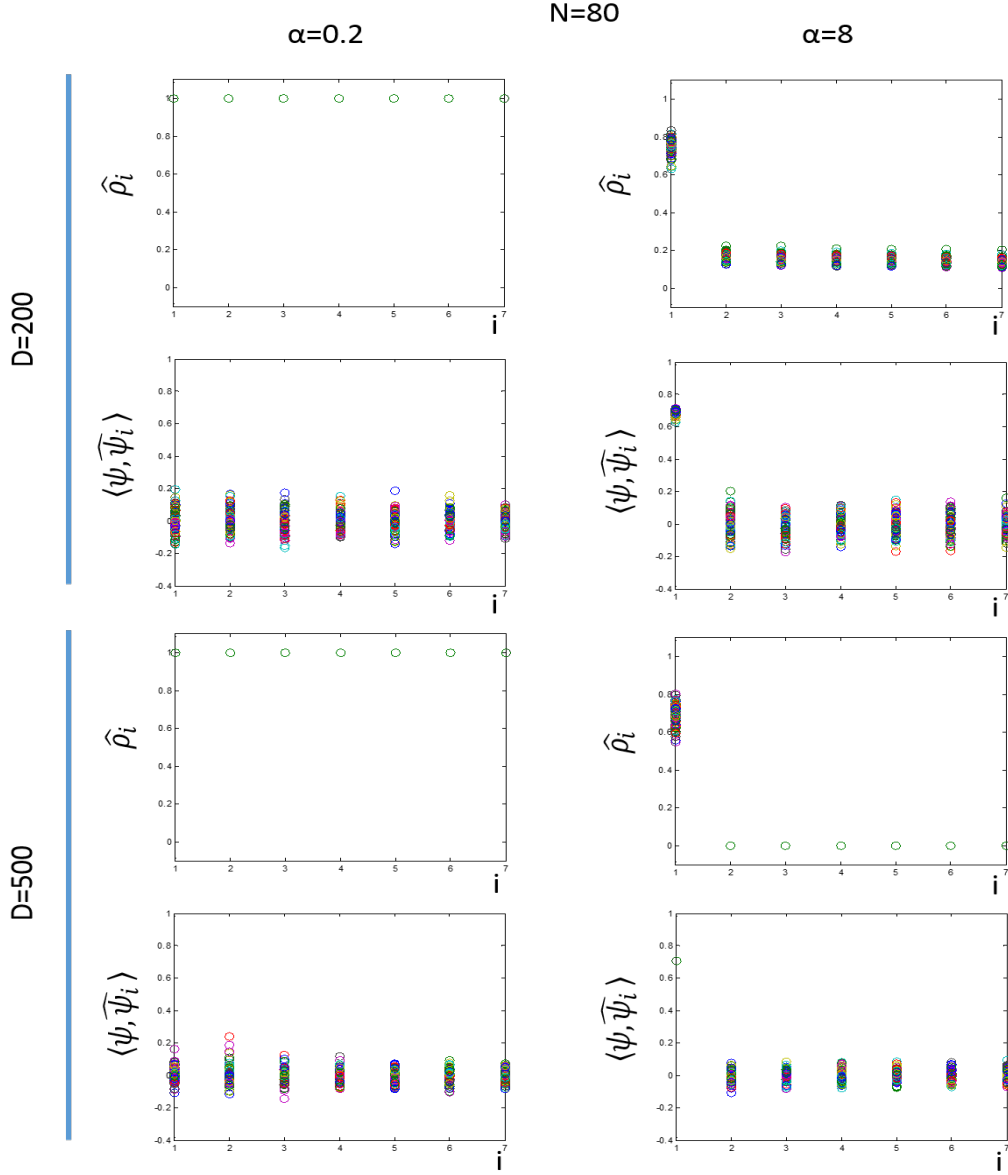


Figure 2. 100 estimated sample canonical correlation coefficients $\hat{\rho}_i^{(d)}$ and inner products of the sample left canonical weight vectors $\hat{\psi}_{X_i}^{(d)}$ and the population canonical weight vector $\psi_{X_i}^{(d)}$, for $i = 1, 2, \dots, 5$, obtained from 100 repetitions of simulations for different settings of dimension d and exponent α with a sample size of $n = 80$.

defined as in (9). We perform 100 runs of simulations for each combination of different values of the following three sets,

- Sample size $n \in \{20, 80\}$,
- Dimension $d \in \{200, 500\}$,

- Exponent $\alpha \in \{0.2, 8\}$.

Each case, estimates of the first 5 canonical correlation coefficients $\hat{\rho}_i^{(d)}$ and their corresponding canonical weight vectors $\hat{\psi}_{Xi}^{(d)}$ and $\hat{\psi}_{Yi}^{(d)}$ are obtained. The estimated vectors $\hat{\psi}_{Xi}^{(d)}$ and $\hat{\psi}_{Yi}^{(d)}$, for $i = 1, 2, \dots, 5$, are compared to the population canonical weight vector $\psi_X^{(d)}$ using their inner product. Here, we do not include results of $\hat{\psi}_{Yi}^{(d)}$ as they are similar as those of $\hat{\psi}_{Xi}^{(d)}$.

Figure 1 presents the simulation results for a small sample size of $n = 20$. For $\alpha = 0.2$, sample coefficients and vectors are almost of no use as the estimated vectors tend to be as far away as possible from the population direction (implied in the inner products of 0) with always perfect correlation. When α increases to a high strength of 8, the first sample coefficient $\hat{\rho}_1^{(d)}$ approaches to the population direction whereas the rest degenerate to 0 as $d \rightarrow \infty$. The first left sample canonical weight vector $\hat{\psi}_{X1}^{(d)}$ converges to the direction $e_1^{(d)}$ (implied in the inner products of $\cos 0.75\pi$) containing dominant variability as $d \rightarrow \infty$ and the rest carry no information on the population direction with tending to deviate from it by a highest degree of 90° . Figure 2 illustrates the results for a larger sample size of $n = 80$. For the case of $\alpha = 0.2$, the behavior of $\hat{\rho}_i^{(d)}$ and $\hat{\psi}_{Xi}^{(d)}$ is similar as that in a small sample size case. However, for $\alpha = 8$, we see a noticeable decrease in variability of the first sample canonical correlation coefficient $\hat{\rho}_1^{(d)}$ around a true value of 0.7 and of the rest of $\hat{\rho}_i^{(d)}$ around 0. This implies that the usual large sample theory works for $\hat{\rho}_1^{(d)}$. Diminishing variability is also observed for the sample canonical weight vectors $\hat{\psi}_{Xi}^{(d)}$, where the first sample vector $\hat{\psi}_{X1}^{(d)}$ becomes almost identical to the largest variance direction $e_1^{(d)}$ and the rest diverge from the population canonical direction $\psi_X^{(d)}$.

6 Discussion

A natural question arises about what asymptotic behavior of sample canonical weight vectors we can expect at the boundary case of $\alpha = 1$ as $d \rightarrow \infty$ with a sample size fixed at n ? When $\alpha > 1$, we saw that the angle between the first sample canonical weight vector and its population counterpart degenerates to 0 and when $\alpha < 1$, the angle between them diverges by as large as $\pi/2$. We conjecture that, with $\alpha = 1$, the angle formed by the first sample canonical weight vector and its population counterpart converges weakly to some random variable on the support of $[0, \pi/2]$. We leave an investigation of this conjecture for future work.

References

- [1] J. Ahn, J.S. Marron, K. Muller, and Y. Chi. The high dimension, low sample size geometric representation holds under mild conditions. *Biometrika*, 94(3):1–7, 2007.
- [2] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.
- [3] S. Jung and J.S. Marron. Pca consistency in high dimension, low sample size context. *The Annals of Statistics*, 37(6B):4104–4130, 2009.
- [4] S. Jung, A. Sen, and J.S. Marron. Boundary behavior in high dimension, low sample size asymptotics of pca. *Journal of Multivariate Analysis*, 109:190–203, 2012.
- [5] M. Lee. Continuum direction vectors in high dimensional low sample size data. *Dissertation, University of North Carolina at Chapel Hill*, pages 54–87, 2007.
- [6] D. Samarov. The analysis and advanced extensions of canonical correlation analysis. *Dissertation, University of North Carolina at Chapel Hill*, pages 121–176, 2009.
- [7] D. Witten and R. Tibshirani. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical Applications in Genetics and Molecular Biology*, 8(1), 2009.